БЕЛКООПСОЮЗ
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БЕЛОРУССКИЙ ТОРГОВО-ЭКОНОМИЧЕСКИЙ
УНИВЕРСИТЕТ ПОТРЕБИТЕЛЬСКОЙ КООПЕРАЦИИ»

Кафедра информационно-вычислительных систем

ЭКОНОМЕТРИКА И ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ

Пособие
для реализации содержания образовательных программ высшего образования I ступени и переподготовки руководящих работников и специалистов

В трех частях

Часть 3

Гомель 2021

УДК 330.42
ББК 65в631
Э 40

Авторы-составители: Л. П. Авдашкова, канд. физ.-мат. наук, доцент;
М. А. Грибовская, канд. физ.-мат. наук, доцент

Рецензенты: И. И. Луханин, канд. техн. наук, доцент, заведующий кафедрой иностранных языков и межкультурных коммуникаций Гомельского филиала Международного университета «МИТСО»;
Л. А. Воробей, канд. физ.-мат. наук, доцент кафедры ИВС Белорусского торгово-экономического университета потребительской кооперации

Рекомендовано к изданию научно-методическим советом учреждения образования «Белорусский торгово-экономический университет потребительской кооперации». Протокол № 4 от 14 апреля 2020 г.

Э 40 **Эконометрика** и экономико-математические методы и модели : пособие для реализации содержания образовательных программ высшего образования I ступени и переподготовки руководящих работников и специалистов. В 3 ч. Ч. 3 / авт.-сост. : Л. П. Авдашкова, М. А. Грибовская. – Гомель : учреждение образования «Белорусский торгово-экономический университет потребительской кооперации», 2021. – 68 с.
ISBN 978-985-540-552-9

В издании излагаются технология построения и анализ эконометрических моделей. Расчеты проводятся с использованием MS Excel. Предназначено для слушателей системы переподготовки руководящих работников и специалистов, аудиторной и самостоятельной работы студентов экономических специальностей.

УДК 330.42
ББК 65в631

ISBN 978-985-540-572-7 (ч. 3)
ISBN 978-985-540-552-9

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Постоянно усложняющиеся экономические процессы требуют повышения уровня образования современных специалистов по экономике и управлению. Изучение дисциплины «Эконометрика и экономико-математические методы и модели» позволяет использовать моделирование и количественный анализ в экономических исследованиях. Специалист экономического профиля для решения поставленных задач должен владеть научными основами исследования социально-экономических систем, анализа исходных данных, формализации, прогнозирования и принятия оптимальных управленческих решений, используя для этой цели современные технические средства.

Пособие «Эконометрика и экономико-математические методы и модели» предназначено для студентов экономических специальностей для организации аудиторной и самостоятельной работы.

В пособии рассмотрены основные эконометрические вопросы, связанные:

- с построением классических линейных регрессионных моделей (парных и множественных);
- с изучением эконометрических моделей, выраженных системой одновременных уравнений;
- с анализом структуры временных рядов и изучением взаимосвязей между ними;
- с рассмотрением основных ошибок, возникающих при нарушении классических модельных предположений, методикой их диагностики и устранения.

Эконометрическое моделирование предполагает получение модели и ее анализ на качество по определенным статистическим параметрам, которые необходимо найти, используя знания математической статистики. Поэтому вначале предлагается выполнить вычисления параметров с помощью приложения MS Excel, описанные в разделе «Технология вычислений в MS Excel …», а затем провести дальнейший анализ построения и оценки модели с помощью полученных значений параметров. Такой подход при построении занятия позволяет, с одной стороны, вспомнить сведения из математической статистики и применить их для решения конкретной задачи эконометрического моделирования на этапе расчетов, с другой стороны, на этапе анализа акцентировать внимание на необходимость выполнения определенных статистических условий, не отвлекаясь на выполнение вычислений.

По каждой теме в пособии предлагается технология вычислений, анализ по полученным параметрам соответствующей модели, вопросы для самоконтроля, индивидуальные задания. Так как этапы эконометрического моделирования справедливы для любой модели, то в темах 1–3 в разделах по технологии вычислений и анализу модели используется одинаковая нумерация этих этапов.

В разделе «Эконометрический анализ построения модели …» для каждого этапа эконометрического моделирования приводится соответствующий теоретический материал, который позволяет организовать обсуждение проблемных ситуаций, улучшить организацию самостоятельной работы, ответить на вопросы для самоконтроля.

Индивидуальные задания позволяют организовать самостоятельную работу студентов.

В приложениях приводятся примеры оформления листов рабочих книг MS Excel, содержащих вычисления, которые описаны в разделе «Технология вычислений …», а также образец оформления отчетов по лабораторным работам.

ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ

Экономические исследования требуют от экономистов умений применять экономико-математические методы, создавать эконометрические модели, основываясь на знании экономической теории, экономической статистики, математического моделирования, теории вероятностей, математической статистики, для понимания количественных взаимосвязей между экономическими факторами в целях анализа и прогнозирования реальных экономических процессов.

Экономическая теория с помощью качественного анализа устанавливает совокупность факторов и показателей, влияющих на изучаемое экономическое явление, их роль и теоретические взаимосвязи. Экономическая статистика обеспечивает информационную базу экономических исследований, осуществляя первичную обработку эмпирических значений выбранных экономических показателей. Экономическая статистика, как правило, ограничивается простейшими качественными выводами. Математическая статистика обеспечивает инструментом работы со случайными величинами. Математическое моделирование формализует рассматриваемую экономическую задачу на языке математики. Эконометрика оценивает количественные взаимосвязи изучаемых факторов и использует эти оценки для прогнозирования экономических процессов.

Одной из основных задач эконометрики является построение и анализ эконометрической модели. При этом под *эконометрической моделью* понимается такая форма представления исследуемой экономической задачи с помощью математических терминов и соотношений на основе статистических данных, которая удобна для проведения количественного анализа.

Существуют различные классификации эконометрических моделей. Например, одна из них выделяет типы эконометрических моделей по фактору времени (*статические* и *динамические*, т. е. *модели временных рядов*). Первые из них исследуют состояние системы в определенный момент времени, т. е. опираются на единовременный срез информации по изучаемым объектам. Вторые строятся по данным, характеризующим изучаемые объекты за ряд последовательных периодов времени, т. е. они используют не только текущие значения показателей, но и некоторые предыдущие по времени значения, а также и само время t.

В зависимости от формы математического представления эконометрические модели подразделяют на *модели с одним уравнением* и *модели системы одновременных уравнений*.

В первом случае объясняемый фактор y выражается через объясняющие переменные $x_1, x_2, ..., x_m$ с помощью одного уравнения, в котором каждому конкретному набору ($x_1, x_2, ..., x_m$) объясняющих факторов $x_1, x_2, ..., x_m$ соответствует некоторое вероятностное значение зависимой переменной y, выраженное математическим ожиданием $M(y | x_1, x_2, ..., x_m)$. Такая зависимость называется *корреляционной*. При этом модель называется *парной*, если уравнение связывает только две переменные x, y и описывается уравнением *парной регрессии* вида $M(y | x) = f(x)$. Если же речь идет о зависимости величины y от нескольких факторов $x_1, x_2, ..., x_m$ ($m \geq 2$), то модель с одним уравнением $M(y | x_1, x_2, ..., x_m) = f(x_1, x_2, ..., x_m)$ называется *множественной моделью*. В зависимости от вида функции эконометрические модели разделяются на *линейные* и *нелинейные*.

Модели системы эконометрических уравнений используются при изучении достаточно сложных экономических явлений взаимосвязи между исследуемыми показателями, которые описываются не одним, а несколькими уравнениями (например, модель равновесия спроса и предложения в рыночной экономике).

Выделяют следующие этапы решения эконометрической задачи:

- *Постановочный этап*, предполагающий определение целей и задач исследования; выделение факторов и показателей, определяющих изучаемые экономические процессы; установление на базе экономической теории роли выбранных показателей.
- *Этап спецификации*, во время которого выбирается формула связи между переменными, обозначающими выделенные факторы. Эта формула имеет общий вид и содержит параметры (коэффициенты при переменных), требующие статистической оценки.
- *Этап параметризации*, решающий задачу оценки значений параметров выбранной функции связи.
- *Этап верификации*, предполагающий проверку адекватности модели, т. е. проверку соответствия модели реальному экономическому явлению или процессу. Кроме того, на данном этапе выясняется, насколько удачно решены проблемы спецификации и параметризации, совершенствуется форма модели, уточняется состав объясняющих переменных, устанавливается точность расчетов по данной модели, определяются общее качество уравнения, статистическая значимость найденных параметров, а также разрешаются многие другие вопросы, определяющие надежность выводов по модели.

Реальное значение зависимой переменной y не совпадает с условным математическим ожиданием $M(y \mid x_1, x_2, ..., x_m)$ и отличается от него на некоторое значение ε, которое носит случайный характер, т. е. $y = M(y \mid x_1, x_2, ..., x_m) + \varepsilon$.

Таким образом, величина y разбивается на две части: одна из них (объясняемая) задает ту часть y, которая объясняется факторами $x_1, x_2, ..., x_m$, вторая часть ε является случайной величиной и определяет влияние на y неучтенных уравнением $M(y \mid x_1, x_2, ..., x_m) = f(x_1, x_2, ..., x_m)$ других факторов.

Уравнение $y = f(x_1, x_2, ..., x_m) + \varepsilon$ называется *регрессионной моделью* (или *уравнением регрессионной модели*).

Общая задача эконометрического моделирования заключается в следующем: по имеющимся данным n наблюдений за изменением признака y в зависимости от наборов значений факторов $x_1, x_2, ..., x_m$ выбрать эконометрическую модель $y = f(x_1, x_2, ..., x_m) + \varepsilon$, оценить ее параметры и статистически обосновать, что факторы $x_1, x_2, ..., x_m$ существенны, а построенная функция $f(x_1, x_2, ..., x_m)$ такова, что наиболее точно соответствует данным наблюдений.

Тема 1. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Постановка задачи

Исследуйте зависимость фактора y от факторов x_1 и x_2, используя данные наблюдений, приведенные в таблице 1. Постройте регрессионную модель $y = f(x_1, x_2) + \varepsilon$. Рассчитайте значение показателя y при $x_1 = 35$ и $x_2 = 10$.

Таблица 1 – **Данные наблюдений**

Фактор		
y	x_1	x_2
684,48	28	10
674,45	26	8
729,62	30	14
748,86	35	15
761,44	41	16
773,42	45	17
628,07	27	3
731,84	35	13
698,81	30	10
645,92	23	5
664,64	29	7
711,18	33	11
798,07	40	20
833,82	41	24
667,97	41	6
607,61	23	2
711,76	32	12
728,62	37	13
666,15	31	7
683,70	30	9

Технология вычислений в MS Excel для построения и анализа линейной множественной регрессии

1. Постановочный этап: ***определение целей и задач исследования; выделение факторов и показателей, определяющих изучаемые экономические процессы; установление роли выбранных показателей*** (дано в условии задачи или определяется исходя из экономической теории); ***подготовка данных для расчетов***

Подготовьте данные для расчетов (введите исходные данные, представленные в таблице 1) следующим образом: в ячейку A1 введите название первого столбца – «Фактор *y*», в ячейку B1 – название второго столбца – «Фактор *x*1», в ячейку C1 – название третьего столбца – «Фактор *x*2». В ячейки A2, A3, …, A21 введите данные первого столбца таблицы 1, в ячейки B2, B3, …, B21 – данные второго столбца, в ячейки C2, C3, …, C21 – данные третьего столбца этой таблицы.

Введите новое название листа «Исходные данные» и сохраните рабочую книгу (*Файл → Сохранить как → …*).

Примечания:
1. В приложении А приведен пример оформления вычислений на листах MS Excel.
2. На листах MS Excel переменные x_1, x_2 будем обозначать *x*1, *x*2.

2. Спецификация: ***выбор в общем виде формулы связи между переменными, обозначающими выделенные факторы***

Вид и сила функциональной зависимости (линейная или нелинейная) определяются по коэффициенту корреляции.

На вкладке *Данные* выберите команду *Анализ данных*.

Примечание – Если в ленте на вкладке Данные отсутствует эта команда, то следует выбрать *Файл → Параметры → Надстройки*, выделить надстройку Пакет анализа, нажать внизу окна кнопку Перейти. В открывшемся окне установить флажок *Пакет анализа* и нажать кнопку ОК.

На листе «Исходные данные» выполните действия *Данные → Анализ данных → Корреляция → ОК*. Значения параметров окна установите следующим образом:

- *Входной интервал* – введите ссылки на ячейки A1:C21.
- *Метки* – установите флажок.
- *Параметры вывода* – установите переключатель на *Выходной интервал* и в поле установите ссылку на ячейку E2. Нажмите кнопку *ОК*. Скопируйте ячейку G5 в H4. В ячейку E1 введите название «Корреляционная матрица».

Для проверки гипотезы о значимости коэффициента корреляции сравниваются наблюдаемое $t = \dfrac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$ и критическое $t_{\alpha,\nu} = t_{кр}$ значения статистики Стьюдента, для нахождения которых выполните следующие действия:

• В ячейку E6 введите название «Значимость коэффициентов корреляции».

• В ячейку E7 введите обозначение tнабл y, $x1$.

• В ячейку F7 для вычисления tнабл для коэффициента корреляции факторовy, $x1$ введите формулу

= F4*КОРЕНЬ((20 – 2)/(1 – F4^2)), где 20 – число наблюдений, 2 – число факторов.

• В ячейку E8 введите tнабл y, $x2$.

• В ячейку F8 для вычисления tнабл для коэффициента корреляции факторов y, $x2$ введите формулу

= F5*КОРЕНЬ((20 – 2)/(1 – F5^2)), где 20 – число наблюдений, 2 – число факторов.

• В ячейку E9 введите tкр.

• В ячейке F9 вычислите критическое значение tкр следующим образом:

– нажмите на кнопку f_x (вставка функций);

– в поле *Категория* окна *Мастер функций* выберите *статистические*, из предложенных ниже функций выделите СТЬЮДЕНТ.ОБР.2Х и нажмите кнопку *ОК*. Откроется окно *Аргументы функций.* Заполните поля:

♦ *Вероятность* – наберите значение 0,05;

♦ *Степени свободы* – введите 20 – 2, где 20 – число наблюдений, 2 – число факторов. Нажмите кнопку *ОК*.

Примечание – Выводы о существовании зависимости и выборе вида функции связи подробно описаны в разделе «Эконометрический анализ построения модели множественной регрессии».

3. Параметризация модели: *нахождение оценок значений параметров выбранной функции связи*

Найдите МНК-оценки неизвестных параметров b_0, b_1 парной линейной регрессионной модели $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$, где ε – случайная переменная, которая включает в себя суммарное влияние всех неучтенных в модели факторов, выполнив действия, представленные ниже.

Примечание – Предполагаем, что между факторами существует линейная зависимость. Далее находится уравнение линейной регрессии. Если доказана нелинейность зависимости, то проводится процедура линеаризации [6].

Выполните действия *Данные → Анализ данных → Регрессия → ОК*. Значения параметров в появившемся окне установите следующим образом:

- *Входной интервал Y* – введите ссылки на ячейки A1:A21.
- *Входной интервал X* – введите ссылки на ячейки B1:C21.
- *Метки* – установите флажок.
- *Уровень надежности* – установите флажок.
- *Константа ноль* – оставьте пустым.
- *Параметры вывода* – установите переключатель на *Новый рабочий лист* и в соответствующее поле введите его название «Регрессия».
- *Остатки* – установите флажок.
- *Стандартизированные остатки* – оставьте пустым.
- *График остатков* – установите флажок.
- *График подбора* – установите флажок.
- *График нормальной вероятности* – оставьте пустым. Нажмите кнопку *ОК*.

Расположите диаграммы рядом (на поле диаграммы нажмите левую кнопку мышки, затем поместите курсор на белое поле и при нажатой левой кнопке передвигайте диаграмму вниз) и растяните (на поле диаграммы нажмите левую кнопку мышки, нижнюю линию границы диаграммы при нажатой левой клавише протяните вниз).

Примечание – Выводы о значениях оценок параметров уравнения регрессии подробно описаны в разделе «Эконометрический анализ построения модели множественной регрессии».

4. Верификация модели: *проверка адекватности модели*

4.1. Общее качество уравнения: *проверка значимости коэффициента детерминации*

Для проверки гипотезы о значимости коэффициента детерминации сравниваются наблюдаемое значение статистики Фишера, найденное с помощью анализа *Регрессия*, и критическое значение $F_{\alpha,\nu_1,\nu_2} = F_{кр}$, которое вычислите на листе «Регрессия» в свободной ячейке E15 следующим образом:

- Нажмите на кнопку f_x (вставка функций);
- В поле *Категория* окна *Мастер функций* выберите *статистические*, из предложенных ниже функций выделите F.ОБР.ПХ и нажмите кнопку *ОК*. Откроется окно *Аргументы функций*. Заполните поля:

– *Вероятность* – наберите значение 0,05;

– *Степени свободы1* – установите курсор в поле и выделите ячейку B12 столбца *df* таблицы «Дисперсионный анализ»;

– *Степени свободы2* – установите курсор в поле и выделите ячейку B13 столбца *df* таблицы «Дисперсионный анализ». Нажмите кнопку *ОК*.В ячейку D15 листа «Регрессия» введите Fкр.

Примечание – Выводы о качестве уравнения подробно описаны в разделе «Эконометрический анализ построения модели множественной регрессии».

4.2. Нормальность распределения остатков: *устанавливается для возможности использования статистики Стьюдента при проверке гипотез (визуально по гистограмме, асимметрии и эксцессу, с помощью проверки параметрической гипотезы)*

На листе «Регрессия» выполните действия *Данные → Анализ данных → Описательная статистика → ОК*. Значения параметров в диалоговом окне установите следующим образом:

• *Входной интервал* – введите ссылки на ячейки C25:C45 (ячейки со значениями остатков и названием «Остатки»).

• *Группирование по столбцам* – установите переключатель.

• *Метки в первой строке* – установите флажок.

• *Параметры вывода* – установите переключатель на *Выходной интервал*, в поле напротив выделите указателем мыши ячейку D26.

• *Итоговая статистика* – установите флажок.

Параметры *Уровень надежности* (95%), *К-ый наибольший, К-ый наименьший* оставьте пустыми. Нажмите кнопку *ОК*.

Для проверки гипотезы о нормальности распределения остатков с помощью критерия хи-квадрат Пирсона сравниваются наблюдаемое и критическое значения статистики хи-квадрат.

Рассчитайте наблюдаемое значение хи-квадрат статистики $\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n \cdot p_i\right)^2}{n \cdot p_i}$, выполнив действия, приведенные ниже.

В ячейку A49 введите название «Критерий Пирсона». Выполните действия *Данные → Анализ данных → Гистограмма → ОК*. Значения параметров окна установите следующим образом:

• *Входной интервал* – введите ссылки на ячейки C25:C45 (ячейки со значениями остатков и названием «Остатки»).

• *Интервал карманов* – не заполняйте.

• *Метки* – установите флажок.

• *Параметры вывода* – установите переключатель на *Выходной интервал* и введите ссылку на ячейку A50.

- *Парето* – оставьте пустым.
- *Интегральный процент* – оставьте пустым.
- *Вывод графика* – установите флажок. Нажмите кнопку *ОК*.

Перенесите гистограмму вниз и растяните ее.

Удалите слово *Еще* в столбце «Карман» и в этой же ячейке введите формулу = E38*3, т. е. значение максимума остатков увеличили в три раза.

В ячейку C50 введите значение 0.

В ячейки C51:C55 введите *формулу массива*:

1) выделите ячейки C51:C55;

2) нажмите функциональную клавишу *F2*;

3) введите формулу = НОРМ.РАСП(A51:A55;E28;E32;ИСТИНА);

4) нажмите комбинацию клавиш *Ctrl+Shift+Enter*.

Если появилось только одно значение, то нажмитеклавишу *F2* и снова *Ctrl+Shift+Enter*. В строке формул при активизации любой ячейки диапазона C51:C55 появится формула в фигурных скобках:

{= НОРМ.РАСП(A51:A55;E28;E32;ИСТИНА)}.

Примечание – В дальнейшем фраза «введите формулу массива» предполагает выполнение четырех действий: 1) выделить заполняемый диапазон ячеек; 2) нажать на клавиатуре клавишу*F2*; 3) ввести формулу; 4) нажать комбинацию клавиш *Ctrl+Shift+Enter*.

В ячейки D51:D55 введите *формулу массива*

{= C51:C55 – C50:C54}.

В ячейки E51:E55 введите *формулу массива* {= E40*D51:D55}.

В ячейки F51:F55 введите *формулу массива*

{= (B51:B55 – E51:E55)^2/E51:E55}.

В ячейку A57 введите обозначение «хи-кв набл».

В ячейку B57 введите формулу = СУММ(F51:F55) для вычисления хи-квадрат набл.

Найдите критическое значение статистики Пирсона.

В ячейку A58 введите обозначение «хи-кв кр».

В ячейку B58 введите формулу = ХИ2.ОБР.ПХ (0,05; 6 – 2 – 1), где 6 = 5 + 1 (сумма числа значений в «кармане» 5 и 1), 2 – число параметров нормального распределения (для вычисления хи-квадрат кр).

Примечание – Выводы о нормальности распределения остатков подробно описаны в разделе «Эконометрический анализ построения модели множественной регрессии».

4.3. Значимость коэффициентов регрессии: *проверка соответствующих гипотез*

Для проверки гипотез о значимости коэффициентов регрессии сравниваются наблюдаемые значения *t*-статистики, найденные с помощью анализа *Регрессия*, и критическое значение, которое необходимо найти нижеописанным способом.

В ячейку C20 листа «Регрессия» введите tкр. Вычислите критическое значение tкр в свободной ячейке D20 следующим образом:

• Нажмите на кнопку f_x (вставка функций).

• В поле *Категория* окна *Мастер функций* выберите *статистические*, из предложенных ниже функций выделите СТЬЮДЕНТ.ОБР.2Х и нажмите кнопку *ОК*. Откроется окно *Аргументы функций*. Значения параметров следующие:

– *Вероятность* – наберите значение 0,05;

– *Степени свободы* – введите 20 – 2 – 1, где 20 – число наблюдений, 2 – число факторов в уравнении регрессии, 1 – число свободных членов (b_0) в уравнении регрессии. Нажмите кнопку *ОК*.

Примечание – Выводы о значимости коэффициентов регрессии подробно описаны в разделе «Эконометрический анализ построения модели множественной регрессии».

4.4. Проверка статистических свойств остатков (качества оценок коэффициентов регрессии)

4.4.1. Центрированность остатков

Для проверки гипотезы о значимости математического ожидания случайной переменной ε, выборочными оценками которой являются остатки, сравниваются наблюдаемое $t = \dfrac{(\varepsilon - 0)\sqrt{n}}{S}$ и критическое $t_{\alpha,\nu} = t_{кр}$ значения *t*-статистики.

На листе «Регрессия» в ячейку D25 введите название «Условие 1».

В ячейку D41 введите *tнабл*. В ячейку E41 введите формулу = (E28 – 0)*КОРЕНЬ(E40)/E32 для подсчета наблюдаемого значения статистики tнабл. В ячейку D42 введите *tкр*. В ячейку E42 введите формулу = СТЬЮДЕНТ.ОБР.2Х(0,05;20 – 1) для подсчета критической точки распределения Стьюдента *tкр*.

4.4.2. Гомоскедастичность (гетероскедастичность) остатков

Для проверки гипотезы о гомоскедастичности случайной переменной сравниваются наблюдаемое и критическое значения t-статистики.

Найдите наблюдаемое значение t-статистики по формуле

$$t = r\sqrt{n-1},$$

где $r = 1 - \dfrac{6\sum D_i^2}{n(n^2-1)}$ – коэффициент ранговой корреляции Спирмена;

D_i – разность между рангом x_i и рангом модуля остатка e_i.

На листе «Исходные данные» содержимое ячеек A1:A21 скопируйте в ячейку A1 нового листа и назовите лист «Условие2». В ячейку B1 скопируйте из листа «Регрессия» столбец «Остатки» вместе с названием. В ячейку C1 введите название «Модуль ост». В ячейки C2:C21 введите *формулу массива*{= ABS(B2:B21)}(выделите ячейки C2:C21, нажмите клавишу*F2*, введите формулу, нажмите комбинацию клавиш *Ctrl*+*Shift*+*Enter*).

Выполните команду *Данные* → *Анализ данных* → *Ранг и персентиль* → *ОК* и заполните диалоговое окно следующим образом:

- *Входной интервал* – введите ссылки на ячейки A1:A21;
- *Метки* – установите флажок;
- *Выходной интервал* – ячейка D1. Нажмите кнопку *ОК*.

Выполните команду *Данные* → *Анализ данных* → *Ранг и персентиль* → *ОК* и заполните диалоговое окно следующим образом:

- *Входной интервал* – введите ссылки на ячейки C1:C21;
- *Метки* – установите флажок;
- *Выходной интервал* – ячейка H1. Нажмите кнопку *ОК*.

Выделите ячейки D2:G21 и нажмите кнопку *Сортировка по возрастанию* на панели инструментов. Выделите ячейки H2:K21 и нажмите кнопку *Сортировка по возрастанию*. В ячейки L2:L21 введите *формулу массива* {= (F2:F21 – J2:J21)^2}. В ячейку L1 введите название «Квадрат разности рангов».

В ячейку K22 введите *tнабл*. В ячейку L22 введите формулу = (1 – 6*СУММ(L2:L21)/(20*(20^2 – 1)))*КОРЕНЬ(20 – 1) для вычисления *tнабл*.

В ячейку K23 введите *tкр*. В ячейку L23 введите формулу = СТЬЮДЕНТ.ОБР.2Х(0,05;20 – 2) для вычисления *tкр*.

4.4.3. Автокорреляция остатков

Для проверки гипотезы об отсутствии автокорреляции случайной переменной сравниваются наблюдаемое и критические значения статистики Дарбина–Уотсона.

Найдите наблюдаемое значение статистики $d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^{\,2}}$, используя в качестве оценок значений случайной переменной соответствующие значения остатков, выполнив действия, приведенные ниже.

На листе «Регрессия» в ячейку F25 введите название «Условие3».

В ячейки F26:F44 введите *формулу массива*

{= (C26:C44 – C27:C45)^2}.

В ячейку F46 введите формулу = СУММ(F26:F44).

В ячейки G26:G45 введите *формулу массива* {= (C26:C45)^2}.

В ячейку G46 введите формулу = СУММ(G26:G45).

В ячейку F47 введите *dнабл*.

В ячейку G47 введите формулу = F46/G46 для вычисления *dнабл*.

Если критерий Дарбина–Уотсона не дает ответа о наличии автокорреляции, то можно воспользоваться визуальным способом анализа графика зависимости остатков от номера наблюдения, построенного с помощью диаграммы.

Для построения графика остатков выполните следующие действия:

- выделите диапазон ячеек C25:C45, содержащий остатки вместе с названием;
- на вкладке *Вставка ленты инструментов* выберите диаграмму типа *Точечная*.

Примечание – Выводы о выполнении условий Гаусса–Маркова подробно описаны в разделе «Эконометрический анализ построения модели множественной регрессии».

4.5. Анализ свойств модели

4.5.1. Мультиколлинеарность факторов: *выявление зависимости объясняющих факторов*

Для проверки гипотезы об отсутствии мультиколлинеарности используется статистика хи-квадрат с $\nu = \frac{n(n-1)}{2}$ степенями свободы, наблюдаемое значение которой определяется по формуле

$$\chi^2 = n - 1 - \frac{1}{6}(2p + 5)\lg \Delta r,$$

где n – количество наблюдений;

p – число независимых переменных.

Δr – определитель матрицы парных коэффициентов корреляции между факторами.

На листе «Исходные данные» в ячейку E11 введите название «Мультиколлинеарность».

В ячейку E12 введите название «Определитель».

В ячейку F12 введите математическую формулу

= МОПРЕД(G4:H5).

В ячейку E13 введите название «хи-кв набл».

В ячейку F13 введите формулу = 20 – 1 – 9*LOG(F12;10)/6 для нахождения хи-квадрат, наблюдаемого по выборке.

В ячейку E14 введите название «хи-кв кр».

В ячейку F14 введите формулу = ХИ2.ОБР.ПХ(0,05;20*(20 – 1)/2) для нахождения хи-квадрат критического.

4.5.2. Эластичность

Средний коэффициент эластичности для i-го фактора линейной регрессии рассчитывается по формуле

$$\overline{Э}_{yx_i} = b_i \frac{\overline{x}_i}{\overline{y}}.$$

На листе «Исходные данные» в ячейку A23 введите название «Эластичность».

В ячейку A24 введите название «Фактор y_ср».

В ячейку A25 введите формулу = СРЗНАЧ(A2:A21).

В ячейку B24 введите название «Фактор x1_ср».

В ячейку B25 введите формулу = СРЗНАЧ(B2:B21).

В ячейку C24 введите название «Фактор x2_ср».

В ячейку C25 введите формулу = СРЗНАЧ(C2:C21).

В ячейку A26 введите название «Коэф. фактора x1». Из листа «Регрессия» скопируйте коэффициент при переменной «Фактор x1» из ячейки B18 в ячейку B26 листа «Исходные данные».

В ячейку A27 введите название «Эластичность фактора x1».

В ячейку В27 введите формулу = B26*B25/A25 для вычисления средней частной эластичности по переменной «Фактор x1».

В ячейку А28 введите название «Коэф. фактора x2». Из листа «Регрессия» скопируйте коэффициент при переменной «Фактор x2» из ячейки В19 в ячейку В28 листа «Исходные данные».

В ячейку А29 введите название «Эластичность фактора x2».

В ячейку В29 введите формулу = B28*C25/A25 для вычисления средней частной эластичности по переменной «Фактор x2».

4.5.3. Частные коэффициенты корреляции

Коэффициент частной корреляции первого порядка для переменной x_1 при неизменном значении переменной x_2 находится по формуле $r_{yx_1 \cdot x_2} = \dfrac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}}$ (через коэффициенты парной корреляции факторов). Для его нахождения выполните следующие действия:

- в ячейку E16 введите название «Частные коэф. корр.»;
- в ячейку E17 введите название «ry,x1-x2»;
- в ячейку F17 введите формулу

= (F4 – F5*G5)/КОРЕНЬ((1 – F5^2)*(1 – G5^2)).

Аналогично найдите $r_{yx_2 \cdot x_1}$:

- в ячейку E18 введите название «ry,x2-x1»;
- в ячейку F18 введите формулу

= (F5 – F4*G5)/КОРЕНЬ((1 – F4^2)*(1 – G5^2)).

Проверка значимости частных коэффициентов осуществляется сравнением наблюдаемых и критического значений *t*-статистики аналогично проверке значимости парных коэффициентов корреляции на этапе спецификации (приложение).

5. Прогнозирование

Точечный прогноз y^* находится подстановкой значений объясняющих переменных 35, 10 в уравнение регрессии.

На листе «Регрессия» в ячейке Е1 введите название «Точечный прогноз», в ячейку Е2 введите формулу = B17 + B18*35 + B19*10 для расчета точечной оценки параметра y при значениях 35 и 10 объясняющих факторов из условия задачи.

Интервальный прогноз, или доверительный интервал прогноза, имеет следующий вид:

$$(y^* - t_{\alpha,\nu}S^*, y^* + t_{\alpha,\nu}S^*),$$

где $t_{\alpha,\nu}$ – критическое значение t-статистики при заданном уровне значимости α и числе степеней свободы ν;

S^* – средняя стандартная ошибка прогноза.

Средняя стандартная ошибка прогноза вычисляется по формуле

$$S^* = S \cdot \sqrt{X_p^T (X^T X)^{-1} X_p},$$

где X – матрица наблюдений независимых переменных;

Xp – матрица значений независимых переменных для прогноза;

S – стандартная ошибка регрессии;

T – операция транспонирования матрицы.

В ячейку B2 нового листа «Интервальный прогноз» скопируйте ячейки B2:C21 листа «Исходные данные». Заполните ячейки A2:A21 единицами (это значения переменной при свободном члене). Для простоты дальнейших ссылок в объединенные ячейки A1:C1 введите название «Массив 1» (массив X, содержащий значения переменной при свободном члене, фактора x1, фактора x2, – ячейки A2:C21), в ячейку D1 – название «Массив 2» (массив Xp, содержащий данные для прогноза, – ячейки D2:D4). В ячейку D2 введите 1, D3 – 35, D4 – 10.

Пример оформления промежуточных вычислений стандартной ошибки прогноза и интервального прогноза приведен на рисунке 1.

Для транспонирования массива 2 введите в ячейки A23:C23 формулу массива {= ТРАНСП(D2:D4)}.

Для транспонирования массива 1 введите в ячейки A25:T27 формулу массива {= ТРАНСП(A2:C21)}.

Результатом произведения транспонированного массива 1 размерностью 3 на 20 и массива 1 размерностью 20 на 3 является массив 3 размерностью 3 на 3, поэтому в ячейки A29:C31 введите формулу массива {= МУМНОЖ(A25:T27;A2:C21)}.

Результатом вычисления обратной матрицы полученного массива 3 является матрица размерностью 3 на 3, которая находится в ячейках A33:C35 по формуле массива {= МОБР(A29:C31)} (массив4).

Результатом произведения транспонированного массива2 размерностью 1 на 3 и массива 4 размерностью 3 на 3 является массив 5 размерностью 1 на 3, поэтому в ячейки A37:C37 введите формулу массива {= МУМНОЖ(A23:C23;A33:C35)}.

Результатом произведения массива 5 размерностью 1 на 3 и массива 2 размерностью 3 на 1 является массив 6 размерностью 1 на 1, поэтому в ячейку A39 введите формулу = МУМНОЖ(A37:C37;D2:D4).

Стандартную ошибку прогноза посчитайте в ячейке A41 по формуле = регрессия!B7*КОРЕНЬ(A39).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		Массив 1		Массив 2																
2	1	28	10	1																
3	1	26	8	35																
4	1	30	14	10																
5	1	35	15																	
6	1	41	16																	
7	1	45	17																	
8	1	27	3																	
9	1	35	13																	
10	1	30	10																	
11	1	23	5																	
12	1	29	7																	
13	1	33	11																	
14	1	40	20																	
15	1	41	24																	
16	1	41	6																	
17	1	23	2																	
18	1	32	12																	
19	1	37	13																	
20	1	31	7																	
21	1	30	9																	
22	Массив 2 трансп.																			
23	1	35	10																	
24	Массив 1 трансп																			
25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	28	26	30	35	41	45	27	35	30	23	29	33	40	41	41	23	32	37	31	30
27	10	8	14	15	16	17	3	13	10	5	7	11	20	24	6	2	12	13	7	9
28	Масив 3 (произведение массива 1 трансп. и массива 1)																			
29	20	657	222																	
30	657	22349	7799																	
31	222	7799	3062																	
32	Массив 4 (обратный к массиву 3)																			
33	1,88412	-0,06943	0,04023																	
34	-0,06943	0,00296	-0,00251																	
35	0,04023	-0,00251	0,0038																	
36	Масив 5 (произведение массива 2 трансп. и массива 4)																			
37	-0,14353	0,00912	-0,00957																	
38	Массив 6 (произведение массива 5 и массива 2)																			
39	0,08014																			
40	Стандартная ошибка прогноза																			
41	1,01838																			
42	Интервальный прогноз																			
43	697,383	701,68																		

Условие 2 | Интервальный_прогноз

Рисунок 1 –**Пример оформления вычислений интервальной оценки прогноза**

Интервальный прогноз величины y рассчитайте в ячейках A43, B43 соответственно по следующим формулам:

= 'регрессия'!E2 – 'регрессия!D20*'Интервальный_прогноз'!A41
(для левого конца интервала);

= 'регрессия'!E2 + 'регрессия'!D20*'Интервальный_прогноз'!A41
(для правого конца интервала).

Примечание – Запись 'регрессия'!E2 означает, что ячейка E2 находится на листе «Регрессия». Набор и редактирование формулы осуществляется в строке формул.

Эконометрический анализ построения модели множественной регрессии

1. Постановочный этап

На практике фактор y зависит от многих других факторов. В условии задачи выделены два наиболее значимо влияющих фактора. Возникает задача количественного описания зависимости выбранных экономических показателей уравнением множественной регрессии $y = f(x_1, x_2) + \varepsilon$ на основе 20 наблюдений экономических показателей.

2. Спецификация модели: *определение наличия зависимости фактора y от факторов x_1 и x_2, а также формы этой зависимости*

Корреляционным полем называется множество точек на плоскости с координатами (x_i, y_i), $i = 1, 2,\ldots, n$, n – объем выборки.

Для характеристики вида связи используется *ковариация*, рассчитываемая по формуле $\operatorname{cov}(x, y) = \overline{xy} - \bar{x} \cdot \bar{y}$. Если $\operatorname{cov}(x, y) > 0$, то возрастание x приводит к увеличению y и связь прямая. Если $\operatorname{cov}(x, y) < 0$, то возрастание x приводит к уменьшению y и связь обратная. Если $\operatorname{cov}(x, y) \to 0$, то экономические показатели не связаны.

Тесноту связи и наличие линейной зависимости изучаемых экономических показателей оценивает *коэффициент парной корреляции* r_{xy} $(-1 \le r_{xy} \le 1)$: $r_{xy} = \dfrac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} = \dfrac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y}$.

Коэффициент корреляции необходимо проверить на значимость (значительно ли отличается от нуля), так как он найден по выборочной совокупности, что может привести к неверным выводам о всей генеральной совокупности. Проверка значимости коэффициента корреляции осуществляется с помощью t-статистики: $t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$. Величина t имеет распределение Стьюдента с $\nu = n-2$ степенями свободы. По выборке находится наблюдаемое значение $t_{набл}$ статистики. Если $|t_{набл}| > t_{\alpha,\nu}$, то коэффициент корреляции значим ($t_{\alpha,\nu} = t_{кр}$ – критическая точка распределения Стьюдента, зависящая только от объема выборки).

Качественная оценка тесноты связи между величинами выявляется по шкале Чеддока (таблица 2).

Таблица 2 – **Шкала Чеддока**

Теснота связи	Значение коэффициента корреляции при наличии	
	прямой связи	обратной связи
Слабая	0,1–0,3	(–0,1)–(–0,3)
Умеренная	0,3–0,5	(–0,3)–(–0,5)
Заметная	0,5–0,7	(–0,5)–(–0,7)
Высокая	0,7–0,9	(–0,7)–(–0,9)
Весьма высокая	0,9–0,99	(–0,9)–(–0,99)

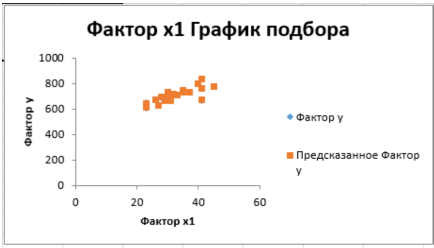
Вид регрессии визуально определяется по корреляционному полю, которое изображено на листе «Регрессия» на графиках подбора черными точками по данным 20 наблюдений из листа «Исходные данные» (рисунок 2).

Поскольку на рисунке 2 точки сгруппированы вдоль прямой (не горизонтальной), то можно предположить, что зависимость фактора y от фактора $x1$линейная и от фактора $x2$ также линейная. Она описывается парной линейной регрессионной моделью

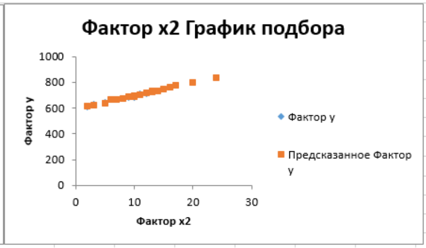
$$y = b_0 + b_1x_1 + b_2x_2 + \varepsilon,$$

где b_0,b_1, b_2– неизвестные параметры модели;

ε – случайная переменная, которая включает в себя суммарное влияние всех неучтенных в модели факторов.



а



б

Условные обозначения:
◆ – фактор *у*;
◇ – предсказанное значение фактора *у*

Рисунок 2 –**Корреляционное поле (график подбора)**

На листе «Исходные данные» получена таблица 3.

Таблица 3 –**Корреляционная матрица**

Корреляционная матрица			
	Фактор у	*Фактор х1*	*Фактор х2*
Фактор у	1		
Фактор х1	0,79	1	
Фактор х2	0,99	0,75	1

Коэффициент корреляции факторов у и *х*1 равен 0,79 > 0, поэтому зависимость между ними прямая и высокая. Коэффициент корреляции факторов у и *х*2 равен 0,99 > 0, поэтому зависимость между ними прямая и весьма высокая (см. таблицу3).

Проверим на значимость коэффициенты парной корреляции. На листе «Исходные данные» вычислены наблюдаемые и критическое значения *t*-статистики (таблица 4).

Таблица 4 –**Значимость коэффициентов корреляции**

Значимость коэффициентов корреляции	
tнабл у, х1	5,56
tнабл у, х2	43,78
tкр	2,10

Поскольку |tнабл у,х1| = 5,56 > tкр = 2,1, то коэффициент корреляции значим (значительно отличается от нуля). Следовательно, подтверждается наличие линейной зависимости между факторами *у* и *х*1.

Поскольку |tнабл y,x2| = 43,79 > tкр = 2,1, то коэффициент корреляции значим. Поэтому также подтверждается наличие линейной зависимости между факторами y и $x2$.

Исходя из проведенного анализа можно выдвинуть предположение о том, что зависимость фактора y от x_1 и x_2 описывается следующей линейной регрессионной моделью:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon,$$

где b_0, b_1, b_2 – неизвестные параметры модели;

ε – случайная переменная, которая включает в себя суммарное влияние всех неучтенных в модели факторов, ошибки измерений.

3. Параметризация модели: *нахождение оценки неизвестных параметров модели*

Статистической оценкой параметра называется его приближенное значение, полученное на основе выборочных данных. Для получения точечных оценок параметров уравнения парной линейной регрессии применяют *метод наименьших квадратов* (МНК). В соответствии с МНК минимизируется сумма квадратов разностей между фактическими и расчетными значениями зависимой переменной. Оценки неизвестных параметров находятся из системы нормальных уравнений, полученной методом дифференциального исчисления.

Доверительные интервалы имеют следующий вид: b_0 – –($b_0 - \Delta_{b_0}$; $b_0 + \Delta_{b_0}$), b_i – ($b_i - \Delta_{b_i}$; $b_i + \Delta_{b_i}$), где центр интервала равен точечной оценке, концы интервалов получены прибавлением и вычитанием произведения стандартной ошибки коэффициента на критическое значение t-статистики Стьюдента $t_{\alpha,\nu}$ для заданного уровня значимости α и числа степеней свободы $\nu = n - 2$.

Доверительный интервал с вероятностью 0,95 содержит истинное значение свободного члена уравнения регрессии. Поэтому любое значение из этого интервала может служить оценкой параметра. Если в границы доверительного интервала попадает ноль, т. е. нижняя граница отрицательна, а верхняя – положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.

В результате проведения регрессионного анализа на листе «Регрессия» получены точечные и интервальные оценки неизвестных параметров модели (таблица 5).

Таблица 5 – **Статистика коэффициентов регрессии**

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	570,74	4,94	115,58	4,63E–26	560,32	581,16
Фактор x1	1,03	0,19	5,26	6,39E–05	0,62	1,44
Фактор x2	9,28	0,22	41,85	1,37E–18	8,81	9,74

Точечная оценка параметра b_0 (Y-пересечение) равна 570,74, ее интервальная оценка – (560,32; 581,16).

Точечная оценка параметра b_1 при переменной x_1 равна 1,03, ее интервальная оценка – (0,62; 1,44).

Точечная оценка параметра b_2 при переменной x_2 равна 9,28, ее интервальная оценка – (8,81; 9,74).

Таким образом, уравнение регрессии имеет следующий вид:

$$y = 570{,}74 + 1{,}03\, x_1 + 9{,}28\, x_2.$$

Поскольку любое значение из доверительного интервала может служить оценкой параметра, то уравнение регрессии также может иметь вид

$$y = 568 + 0{,}8\, x_1 + 9\, x_2.$$

4. Верификация модели

4.1.Общее качество уравнения: *оценка общего качества модели*

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации, который представляет собой долю дисперсии, объясненной выбранным фактором:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2},$$

где e_i – остатки, т. е. отклонения наблюдаемых значений зависимой переменной от теоретических. Для парной регрессии коэффициент детерминации R^2 ($0 \le R^2 \le 1$) рассчитывается как квадрат коэффициента корреляции, который в Excel в таблице *Регрессионная статистика* анализа *Регрессия* называется множественный R. Скорректированный (нормированный) индекс детерминации $\overline{R}^2$ позволяет учесть при оценке качества модели соотношение количества наблюдений и количества оцениваемых параметров модели.

$$\overline{R}^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-k-1},$$

где R^2 – коэффициент детерминации, n – общее число наблюдений, k – число объясняющих переменных (число параметров модели регрессии без учета свободного члена).

Скорректированный коэффициент детерминации применяется для решения двух типов задач:

– оценка тесноты связи между объясняемой и объясняющей переменной. Модель считается качественной, если нескорректированный и скорректированный коэффициенты детерминации близки к единице и значимы.

– сравнение моделей с различным числом параметров. При прочих равных условиях, предпочтение отдается той модели, у которой скорректированный коэффициент детерминации больше.

Если $R^2 = 1$, то все точки наблюдения лежат на регрессионной прямой. Для определения статистической значимости коэффициента детерминации используется F-статистика, рассчитываемая по формуле

$$F = \frac{R^2(n-2)}{1-R^2}.$$

Величина F имеет распределение Фишерас $v_1 = 1$, $v_2 = n - 2$ степенями свободы. По выборке находится наблюдаемое значение Fнабл статистики. Если Fнабл $> F_{\alpha, \nu_1, \nu_2}$, то R^2 значим (F_{α, ν_1, ν_2} = Fкр – критическая точка распределения Фишера, зависящая только от объема выборки).

Оценим общее качество модели по коэффициенту (индексу) детерминации и нормированному индексу детерминации. Проанализируем показатели, представленные в таблице «Регрессионная статистика» листа «Регрессия» (таблица 6).

Таблица 6 –**Регрессионная статистика**

Регрессионная статистика	
Множественный R	0,998
R-квадрат	0,996
Нормированный R-квадрат	0,996
Стандартная ошибка	3,597
Наблюдения	20

Коэффициент множественной детерминации R-квадрат равен 0,996. Поскольку он близок к 1, то уравнение имеет высокое качество. Этот факт подтверждает также нормированный индекс множественной детерминации, равный 0,996.

В таблице «Дисперсионный анализ» листа «Регрессия» рассчитаны наблюдаемое и критическое значения критерия Фишера (таблица 7).

Таблица 7 –**Дисперсионный анализ**

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	61 917,59	30 958,8	2 392,35	1,47E–21
Остаток	17	219,99	12,94		
Итого	19	62 137,59			
			Fкр	3,59	

Поскольку наблюдаемое значение Fнабл =2 392,35 > Fкр = 3,59, то R-квадрат значим, что еще раз подтверждает высокое качество построенного уравнения линейной множественной регрессии.

4.2. Нормальность распределения остатков

Проанализируем нормальность распределения остатков для возможности использования критерия Стьюдента при проверке статистических гипотез. Сделать вывод о нормальности распределения остатков можно следующим образом:

- по гистограмме остатков;
- по числовым характеристикам асимметрии и эксцессу;
- по критерию Пирсона.

> Остатки *e* (отклонения наблюдаемых значений от теоретических) являются оценками случайного члена ε уравнения регрессии.

Анализируя качество модели, необходимо проверить ряд статистических гипотез, использующих критерий Стьюдента, которым можно воспользоваться в случае, когда остатки распределены по нормальному закону.

Кривая плотности нормального распределения задается функцией

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

где a – математическое ожидание;

σ – среднее квадратическое отклонение.

Например, при $a = 0$ и $\sigma = 1$ кривая имеет вид, приведенный на рисунке 3.

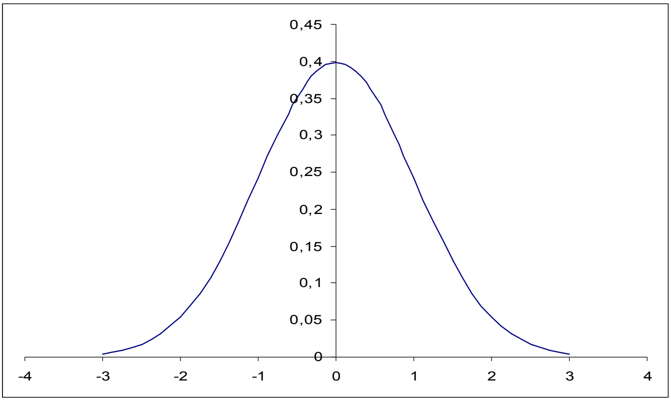
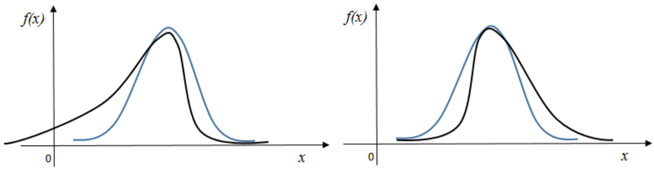


Рисунок 3 – **Кривая нормального распределения**

Визуально нормальность распределения остатков можно определить, сравнивая кривую плотности нормального распределения с гистограммой частот (частостей) остатков, т. е. со ступенчатой фигурой, состоящей из прямоугольников, основаниями которых служат интервалы одинаковой длины на оси Ox, а высоты равны сумме частот значений остатков, попадающих в интервал. Если линия, соединяющая середины верхних сторон прямоугольников, близка к кривой плотности нормального распределения, то предполагают, что распределение остатков приближено к нормальному.

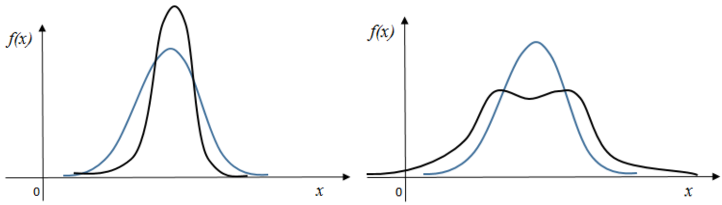
Асимметрия и эксцесс как числовые характеристики нормально распределенной случайной величины равны 0. При асимметричном распределении вершина кривой сдвинута относительно ординаты выборочной средней. Если асимметрия больше 0, то вершина сдвинута вправо (положительная асимметрия), если меньше 0, то – влево (отрицательная асимметрия) (рисунок 4).



(*а*) (*б*)

Рисунок 4 – **Левосторонняя (*а*) и правосторонняя (*б*) асимметрии распределения**

Эксцесс характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением. Положительный эксцесс обозначает относительно остроконечное распределение, отрицательный эксцесс – относительно сглаженное распределение (рисунок 5).



(*а*) (*б*)

Рисунок 5 – **Положительный (*а*) и отрицательный (*б*) эксцессы распределения**

Оценки вида кривой Гаусса и значений асимметрии и эксцесса являются качественными характеристиками распределения. Для надежности вывода (с вероятностью 0,95) проверяется статистическая гипотеза о нормальности распределения с помощью критерия согласия Пирсона. Выдвигается гипотеза о нормальном законе распределения остатков. Для проверки данной гипотезы используется статистика $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$, имеющая распределение χ^2 с $(k - r - 1)$ степенями свободы, где r – число параметров распределения $F(x)$, которые оцениваются по выборке, n – объем выборки; k – число непересекающихся интервалов выборочных значений, n_i – число значений выборки, принадлежащих i-му интервалу, $i = 0, 1, \ldots, k - 1$; p_i – вероятности попадания значений случайной величины в каждый из этих интервалов. По выборке вычисляется наблюдаемое значение статистики $\chi^2_{набл}$. Для выбранного уровня значимости αпо распределению χ^2 находится число $\chi^2_{кр} = \chi^2(\alpha; k - r - 1)$. Гипотеза о нормальном распределении случайного члена принимается на заданном уровне значимости, если $\chi^2_{набл} < \chi^2_{кр}$. Если же $\chi^2_{набл} \geq \chi^2_{кр}$, то гипотеза отвергается.

• Построим *гистограмму остатков*. Соединим середины верхних сторон прямоугольников гистограммы и получим полигон распределения, по которому визуально можно предположить закон распределения (рисунок 6).

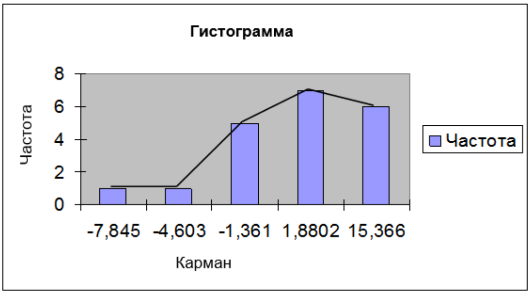


Рисунок 6 – **Гистограмма**

Поскольку ломаная линия на рисунке 6 близка к кривой нормального распределения, заданной уравнением $f(x) = \frac{1}{3{,}4 \cdot \sqrt{2\pi}} e^{-\frac{(x-0)^2}{2 \cdot (3{,}4)^2}}$

(сравните с рисунком 3), то остатки распределены по нормальному закону. Следовательно, по визуальному анализу гистограммы можно предположить нормальность распределения остатков.

• *Асимметричность* равна –0,36 (левосторонняя асимметричность эмпирической кривой относительно теоретической), *эксцесс* равен 0,1 («островершинность» эмпирической кривой), т. е. характеристики плотности распределения (асимметричность и эксцесс) незначительно отличаются от нуля, поэтому можно считать распределение нормальным.

• Подтвердим нормальность распределения с помощью *критерия Пирсона*.

На листе «Регрессия» найдены наблюдаемое и критическое значения статистики хи-квадрат (таблица 8).

Таблица 8 – **Проверка критерия Пирсона**

хи-кв набл	3,16
хи-кв кр	7,81

Наблюдаемое значение, равное 3,16, меньше хи-квадрат критического, равного 7,81, поэтому остатки распределены по нормальному закону.

4.3. Значимость коэффициентов регрессии: *проверка значимости коэффициентов регрессии*

Проверка значимости коэффициентов регрессии осуществляется с помощью *t-статистики Стьюдента.* Вычисляются наблюдаемые значения $t_{b_0} = \frac{b_0}{S_{b_0}}$ и $t_{b_1} = \frac{b_1}{S_{b_1}}$. По таблице критических точек распределения Стьюдента по заданному уровню значимости α и числу степеней свободы $\nu = n - 2$ находится критическая точка $t_{\alpha,\nu} = t_{кр}$. Если $|t_{b_0}| > t_{\alpha,\nu}$, то коэффициент регрессии b_0 значим. Если $|t_{b_1}| > t_{\alpha,\nu}$, то коэффициент регрессии b_1 значим. Значимость коэффициентов подтверждает правильность выбора модели на этапе спецификации. Если хотя бы один из коэффициентов не значим, то необходимо вернуться на этап спецификации.

Значимость коэффициентов регрессии оценивается с помощью t-статистики, значения которой получены на листе «Регрессия» (см. таблицу 5).

Наблюдаемое значение статистики tнабл для коэффициента b_0равно 115,59 (оно равно отношению точечной оценки коэффициента b_0 к его стандартной ошибке). Критическое значение tкр равно 2,1. Поскольку |tнабл| = 115,59 > tкр = 2,1, то коэффициентb_0 значим.

Аналогично для коэффициента b_1 имеем следующее: tнабл = 5,26, tкр = 2,1. Поскольку |tнабл| = 5,26 > tкр = 2,1, поэтому коэффициент b_1 значим. Для коэффициента b_2 имеем: |tнабл| =41,85 > tкр = 2,1, поэтому коэффициент b_2 значим.

Значимость коэффициентов регрессии подтверждает выдвинутое на этапе спецификации предположение о линейной форме зависимости факторов.

4.4. Проверка статистических свойств остатков (качества оценок коэффициентов регрессии)

По статистическим данным нельзя получить точные оценки неизвестных коэффициентов регрессии, но можно найти их приближенные оценки, качество которых характеризуется следующими тремя свойствами:

- ***несмещенностью*** (каждый параметр регрессии можно рассматривать как среднее значение из возможного большого количества оценок);
- ***эффективностью*** (оценки параметров характеризуются наименьшей дисперсией, что в практических исследованиях означает возможность перехода от точечного оценивания к интервальному;
- ***состоятельностью*** (характеризует увеличение точности оценок с увеличением объема выборки).

Метод наименьших квадратов обеспечивает указанные свойства оценкам параметров регрессии при выполнении следующих предположений:

1) математическое ожидание случайной переменной равно нулю;

2) дисперсии наблюдений случайной переменной одинаковы (это условие называется *условием гомоскедастичности*);

3) независимость значений случайной переменной в любом наблюдении от ее значений во всех других наблюдениях.

Предположения 1–3 называются *условиями Гаусса–Маркова.*

Теорема Гаусса–Маркова. *Если условия 1–3 выполнены, то оценки коэффициентов парной линейной регрессии, полученные с помощью МНК, являются несмещенными, состоятельными и эффективными.*

4.4.1. Центрированность остатков: *проверка выполнения условия 1 о равенстве математического ожидания остатков нулю*

Оценкой математического ожидания случайной переменной $M(\varepsilon)$ является *среднее остатков* $\bar{e}$. Гипотеза о значимости математического ожидания случайной переменной проверяется с помощью t-статистики, наблюдаемое значение которой определяется равенством $t = \frac{(\varepsilon - 0)\sqrt{n}}{S}$. По распределению Стьюдента по заданному уровню значимости α и числу степеней свободы $v = n - 1$ находится критическая точка $t_{\alpha,\nu} = t_{кр}$. Если $t_{набл} > t_{кр}$, то $M(\varepsilon)$ значимо. Если $t_{набл} < t_{кр}$, то $M(\varepsilon)$ незначимо.

Если условие 1 нарушено, то оценки коэффициентов регрессии могут быть смещенными, поэтому для устранения этих проблем в модель регрессии следует включать свободный член.

Среднее из числовых характеристик остатков рассчитано на листе «Регрессия» и представлено в таблице 9.

Таблица 9 – **Числовые характеристики остатков**

Условие 1	
Остатки	
Среднее	–2,84E–14
Стандартная ошибка	0,76
Медиана	–0,72
Мода	#Н/Д
Стандартное отклонение	3,40
Дисперсия выборки	11,58
Эксцесс	0,10
Асимметричность	–0,36
Интервал	12,97
Минимум	–7,84
Максимум	5,12
Сумма	–5,68E–13
Счет	20
tнабл	–3,74E–14
tкр	2,09

Среднее остатков равно –2,84E–14 = $-2{,}84 \cdot 10^{-14}$. Оно достаточно близко к нулю, поэтому можно предположить выполнимость условия 1 Гаусса–Маркова. Проверим значение среднего на значимость, т. е. гипотезу о равенстве нулю математического ожидания случайной переменной.

Сравним рассчитанные наблюдаемое и критическое значения статистики. Поскольку |tнабл| = 3,74E–14 = $3{,}74 \cdot 10^{-14} <$ tкр = 2,09, то среднее незначимо (т. е. незначительно отличается от нуля). Следовательно, условие 1 Гаусса–Маркова выполняется.

4.4.2. Гомоскедастичность (гетероскедастичность) остатков: *проверка выполнения условия 2 о постоянстве и конечности дисперсии остатков, т. е. гомоскедастичности остатков*

Предположение о постоянстве и конечности дисперсии остатков называется *свойством гомоскедастичности остатков*. Если оно не выполняется, то такое явление называется *гетероскедастичностью*. Гетероскедастичность часто вызывается ошибками спецификации, когда в модели не учитывается существенная переменная. Гетероскедастичность приводит к тому, что оценки коэффициентов регрессии не являются эффективными, увеличиваются дисперсии распределений оценок коэффициентов, появляется вероятность неверного вычисления оценок стандартных ошибок коэффициентов регрессии. В результате можно сделать неверный вывод о значимости коэффициента. Для оценки нарушения гомоскедастичности наиболее часто используются графический анализ отклонений, тест ранговой корреляции Спирмена и тест Голдфелда–Квандта.

При применении теста Спирмена предполагается, что абсолютные величины остатков и значения объясняющей переменной коррелированны. Эту корреляцию можно измерять с помощью коэффициента ранговой корреляции Спирмена по формуле

$$r = 1 - \frac{6\sum D_i^2}{n(n^2-1)},$$

где D_i– разность между рангом x_i и рангом модуля остатка $e_i = y_i - \tilde{y}_i$.

Тест проводится по следующей схеме:

- Строится линейная модель регрессии.

• Определяются ранги значений x_i независимой переменной и соответствующие ранги модулей остатков $|e_i|$ (ранг – это порядковый номер значения в ранжированном ряду).

• Находится коэффициент ранговой корреляции Спирмена.

• Осуществляется проверка гипотезы об отсутствии гетероскедастичности с помощью t-статистики, наблюдаемое значение которой определяется равенством $t_{набл} = r\sqrt{n-1}$. Если $t_{набл} > t_{кр}$, то гетероскедастичность присутствует. Значит, МНК-оценки неэффективны.

Поскольку $M(\varepsilon)$ равно нулю, то МНК-оценки параметров являются несмещенными и состоятельными, поэтому их позволительно использовать, например, для точечного прогнозирования даже в случае гетероскедастичности. Однако в этом случае МНК-оценки не являются эффективными, а следовательно, результаты (доверительные интервалы для коэффициентов и прогнозных значений), основанные на анализе дисперсии, неверны.

Существуют два подхода к решению проблемы гетероскедастичности:

• преобразование выборочных данных;

• применение взвешенного и обобщенного МНК (ОМНК).

Первый подход предполагает такое преобразование исходных данных, чтобы для них модель уже обладала свойством гомоскедастичности. Используют такие преобразования, как логарифмирование данных, переход к безразмерным величинам путем деления на некоторые известные величины той же размерности, что и исходные данные, стандартизация исходных данных.

Второй подход устранения гетероскедастичности состоит в построении моделей, учитывающих гетероскедастичность ошибок наблюдений.

Обобщенный метод наименьших квадратов (ОМНК) применяется в тех случаях, когда нарушены условия Гаусса-Маркова, касающиеся характера случайных остатков, а именногомоскедастичность (постоянство дисперсии) случайных остатков и некоррелированность остатков между собой.

Нарушение этих условий означает, что ковариационная матрица остатков Ω не является скалярной:на главной диагонали этой матрицы расположены дисперсии случайных остатков для различных наблюдений, не одинаковые по своей величине. Остальные элементы матрицы (ковариации случайных остатков) в общем случае также ненулевые. Если остальные условия построения классической нормальной линейной модели выполняются, то модельназываетсяобобщенной линейной моделью.

Сущность обобщенного метода наименьших квадратов состоит в том, чтобы устранить нарушения предпосылок МНК, «скорректировав» расчеты параметров уравнения регрессии с учетом значений ковариационной матрицы остатков, используя формулу

a’= (XTΩ-1XΩ)-1XTΩ-1Y

где Ω – ковариационная матрица остатков.

С помощью обычного МНК при выполнении всех условий Гаусса-Маркова формула для расчета вектора-столбца неизвестных параметров в матричной форме имеет вид

a= (XTX)-1XTY

На листе «Условие 2» рассчитаны наблюдаемое и критическое значения *t*-статистики (таблица 10).

Таблица 10 – **Проверка гипотезы об отсутствии гетероскедастичности**

tнабл	–1,25
tкр	2,10

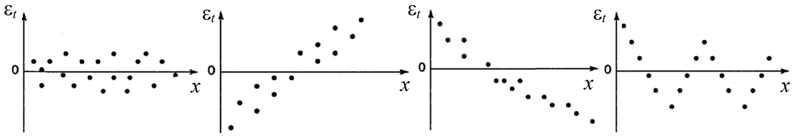
Поскольку |tнабл| = 1,25 < tкр = 2,1, то гетероскедастичность отсутствует. Следовательно, условие 2 Гаусса–Маркова выполняется.Значит, МНК-оценки параметров регрессии будут эффективными. Поэтому модель можно использовать при точечном и интервальном прогнозировании.

4.4.3. Автокорреляция остатков: *проверка выполнения условия 3 о независимости случайного члена в любом наблюдении от его значений во всех других наблюдениях*

Условие 3 Гаусса–Маркова требует *независимости* значений случайной переменной в любом наблюдении от ее значений во всех других наблюдениях. Если данное условие не выполняется, то говорят, что случайная переменная подвержена *автокорреляции*. В этом случае коэффициенты регрессии, получаемые по МНК, оказываются *неэффективными*, хотя и *несмещенными*, а их стандартные ошибки рассчитываются некорректно (занижаются).

Существуют несколько методов определения автокорреляции остатков, два из которых приведены ниже.

Первый метод – это построение графика зависимостей остатков от номера наблюдений и визуальное определение наличия автокорреляции остатков (рисунок 7).



а *б* *в* *г*

Рисунок 7 – **Модели зависимости остатков от номера наблюдения: случайные остатки (*а*); наличие зависимости в остатках (*б*), (*в*), (*г*)**

Второй метод – проверка гипотезы об отсутствии автокорреляции остатков с помощью критерия Дарбина–Уотсона, т. е.

$$d = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} \varepsilon_t^{\ 2}},$$

наблюдаемое значение которого рассчитывается как отношение суммы квадратов разностей последовательных значений остатков к сумме квадратов остатков.

Это же значение критерия может быть вычислено по формуле

$$d = 2(1 - r_1^{\varepsilon}),$$

где $r_1^{\varepsilon} = \dfrac{\sum_{t=2}^{n} (\varepsilon_t - \bar{\varepsilon}_1)(\varepsilon_{t-1} - \bar{\varepsilon}_2)}{\sqrt{\sum_{t=2}^{n} (\varepsilon_t - \bar{\varepsilon}_1)^2 \sum_{t=2}^{n} (\varepsilon_{t-1} - \bar{\varepsilon}_2)^2}}$ – коэффициент автокорреляции первого порядка; $\bar{\varepsilon}_1 = \dfrac{\sum_{t=2}^{n} \varepsilon_t}{n-1}$; $\bar{\varepsilon}_2 = \dfrac{\sum_{t=2}^{n} \varepsilon_{t-1}}{n-1}$.

Альтернативные гипотезы – гипотезы о наличии положительной или отрицательной автокорреляции в остатках.

Если автокорреляция *отсутствует*, то $r_1^{\varepsilon} = 0$ и значение статистики $d \approx 2$. При *положительной* автокорреляции $r_1^{\varepsilon} > 0$ и $0 \leq d < 2$, а при *отрицательной* – $r_1^{\varepsilon} < 0$ и $2 < d \leq 4$. Следовательно, $0 \leq d \leq 4$.

Однако существуют области неопределенности, связанные с тем, что распределение статистики Дарбина–Уотсона зависит не только от числа наблюдений и числа объясняющих переменных, но и от значений объясняющих переменных. В этом случае используются другие способы проверки (например, визуальный и др.).

По таблице, фрагмент которой приведен в таблице 11, определяются критические значения критерия Дарбина–Уотсона d_1 и d_2 для заданного числа наблюдений n, числа объясняющих переменных k и заданного уровня значимости 0,05.

Таблица 11 – **Статистика Дарбина–Уотсона: d_1 и d_2 при уровне значимости 5% (фрагмент таблицы)**

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83

По этим значениям отрезок [0; 4] разбивается на пять областей (рисунок8).

Положительная автокорреляция	Область неопределенности	Автокорреляция отсутствует	Область неопределенности	Отрицательнаяавтокорреляция

$0 \qquad d_1 \qquad d_2 \qquad 4 - d_2 \qquad 4 - d_1 \qquad 4$

Рисунок 8 –**Критические области статистики Дарбина–Уотсона**

В зависимости от того, в какую область попадает наблюдаемое значение критерия, принимают или отвергают гипотезу.

На листе «Регрессия»в ячейке E47 рассчитано наблюдаемое значение d-статистики: dнабл = 1,38. По таблице критических значений d-статистики для числа наблюдений 20, числа объясняющих переменных 2 и заданного уровня значимости 0,05 значения $d_1 = 1{,}10$ и $d_2 = 1{,}54$, которые разбивают отрезок [0; 4] на пять областей (рисунок 9).

Положительная автокорреляция	Область неопределенности	Автокорреляция отсутствует	Область неопределенности	Отрицательная автокорреляция

0 1,10 1,54 2,46 2,90 4

Рисунок 9 –**Критические области *d*-статистики Дарбина–Уотсона**

Поскольку 1,10 < dнабл = 1,38 < 1,54, т. е. наблюдаемое значение попало в зону неопределенности, то ничего нельзя сказать о наличии автокорреляции, используя критерий Дарбина–Уотсона.

Визуально наличие автокорреляции остатков можно определить по графику остатков, полученному на листе «Регрессия» (рисунок 10).

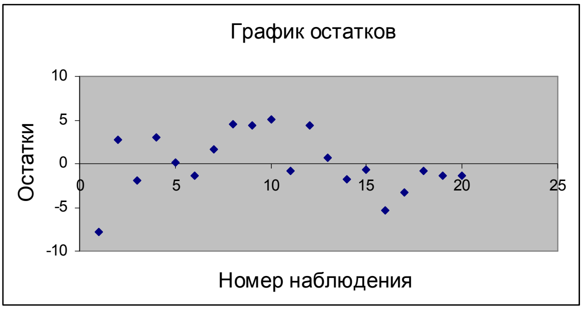


Рисунок 10 – **График остатков**

Поскольку на графике остатков точки разбросаны вдоль прямой $y = 0$ хаотично без видимой закономерности, то зависимости между остатками не наблюдается. Поэтому условие 3 выполняется.

4.5. Анализ свойств модели

4.5.1. Мультиколлинеарность факторов: *проверка мультиколлинеарности факторов*

При построении уравнения множественной регрессии может возникнуть проблема *мультиколлинеарности* факторов, их тесной, линейной связанности. Считается, что две переменные явно *коллинеарны*, т. е. находятся между собой в линейной зависимости, если их парный коэффициент корреляции больше или равен 0,7. При наличии мультиколлинеарности МНК-оценки формально существуют, но обладают рядом недостатков. В частности, оценки имеют большие стандартные ошибки, малую значимость, в то время как модель в целом является значимой с высоким значением коэффициента детерминации.

Для отбора факторов в модель регрессии можно использовать корреляционную матрицу. Однако по величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Поэтому при оценке мультиколлинеарности факторов предполагается использовать определитель Δr матрицы парных коэффициентов корреляции между факторами.

Если бы факторы не коррелировали между собой, то матрица парных коэффициентов корреляции между факторами была бы единичной матрицей, поскольку все недиагональные элементы были бы равны нулю. Если же, наоборот, между факторами существует полная линейная зависимость и все коэффициенты корреляции равны 1, то определитель такой матрицы равен 0, т. е. $\Delta r = 0$. Таким образом, чем ближе к 0 определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И, наоборот, чем ближе к 1 определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Мультиколлинеарность факторов выявляется проверкой гипотезы $H_0 : \Delta r = 1$ с помощью статистики хи-квадрат с $\nu = \frac{1}{2}n(n-1)$ степенями свободы. Наблюдаемое значение статистики определяется по формуле

$$\chi^2_{набл} = n - 1 - \frac{1}{6}(2p + 5)\lg \Delta r,$$

где n – количество наблюдений;

p – число переменных.

Если $\chi^2_{набл} > \chi^2_{\alpha,\nu}$, то гипотеза H_0 отклоняется и наличие мультиколлинеарности объясняющих факторов считается доказанной.

На листе «Исходные данные» найдены парные коэффициенты корреляции и определитель матрицы парных коэффициентов корреляции объясняющих факторов $x1$ и$x2$. Так как парный коэффициент корреляции факторов $x1$ и $x2$ r $x1$, $x2$ = 0,75 < 0,8, то зависимость между факторами существует, но она незначительная. Докажем это предположение проверкой гипотезы об отсутствии мультиколлинеарности с помощью статистики хи-квадрат, наблюдаемое и критическое значения которой найдены на листе «Регрессия» (таблица 12).

Таблица 12 – **Мультиколлинеарность**

Мультиколлинеарность	
Определитель	0,44
хи-кв набл	19,53
хи-кв кр	223,16

Поскольку хи-квадрат наблюдаемое равно 19,53 и меньше хи-квадрат критического, равного 223,16, то мультиколлинеарность факторов отсутствует.

4.5.2. Эластичность: *оценка влияния каждого объясняющего фактора на результирующий фактор y*

Частные средние коэффициенты эластичности показывают, на сколько процентов от среднего значения изменяется зависимая переменная с изменением на 1% фактора x_j от своего среднего при фиксированном значении других факторов. *Частные коэффициенты эластичности* по каждой объясняющей переменной для линейной регрессии рассчитываются по формуле

$$\overline{Э}_{yx_j} = b_j \frac{\overline{x}_j}{\overline{y}}.$$

На листе «Исходные данные» найдены коэффициенты эластичности факторов –x1, x2 (таблица 13).

Таблица 13 –**Эластичность факторов**

Эластичность		
Фактор y_ср	Фактор x1_ср	Фактор x2_ср
707,5215	32,85	11,1
Коэффициент фактора x1	1,03	
Эластичность фактора x1	0,05	
Коэффициент фактора x2	9,28	
Эластичность фактора x2	0,15	

С изменением значения фактора $x1$ на 1% при фиксированном значении фактора $x2$ значение фактора y увеличивается на 0,05%. Аналогично, с изменением значения фактора $x2$ на 1% при фиксированном значении фактора $x1$ значение фактора y увеличивается на 0,15%. Значит, влияние фактора $x2$больше, чем фактора $x1$.

4.5.3. Частные коэффициенты корреляции: *целесообразность включения в модель факторов, определение степени влияния факторов на результирующий фактор y при устранении влияния других факторов*

Частные коэффициенты (или индексы) корреляции характеризуют тесноту связи между результатом и соответствующим фактором при неизменном уровне других факторов, включенных в уравнение регрессии. Они широко используются при решении проблемы отбора факторов, ранжировании факторов, участвующих в множественной линейной регрессии. При нелинейной взаимосвязи исследуемых признаков эту функцию выполняют частные индексы детерминации.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Например, $r_{yx_1 \cdot x_2} = \dfrac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 \cdot x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1 x_2})}}$ – коэффициент частной корреляции первого порядка для переменной x_1при неизменном значении переменной x_2. Аналогично определяется $r_{yx_2 \cdot x_1}$ через коэффициенты парной корреляции факторов.

Значимость частных коэффициентов корреляции оценивается с помощью t-статистики $t = r\sqrt{(n-2)/(1-r^2)}$, которая имеет распределение Стьюдента. Если $|t| \geq t_{\alpha,k}$, то проверяемый коэффициент значим.

На листе «Исходные данные» найдены частные коэффициенты корреляции (таблица 14).

Таблица 14 –**Частные коэффициенты корреляции**

Частные коэффициенты корреляции		Значимость	
r y,x1-x2	0,79	tнабл (r y,x1-x2)	5,41
r y,x2-x1	0,995	tнабл (r y,x2-x1)	43,06

Поскольку 0,78 < 0,995, то из двух факторов большее влияние оказывает фактор $x2$.

Оба частных коэффициента корреляции значимы: |tнабл (r y,x1-x2)| = = 5,4 > tкр = 2,1, |tнабл (r y,x2-x1)| = 43,06 > tкр = 2,1.

***Общий вывод по результатам этапа верификации*:** так как выполняются все условия верификации, то модель является качественной. Таким образом, прогноз, выполненный по ней, является качественным, т. е. несмещенным, состоятельным и эффективным.

5. Прогнозирование

Если выполняются все условия верификации, то модель является качественной. В противном случае ее надо усовершенствовать либо на этапе спецификации, либо варьировать выборку. По качественной модели можно прогнозировать значение зависимой переменной при заданных значениях независимых переменных. Точечный прогноз получается подстановкой заданных значений независимых переменных в уравнение регрессии. Интервальный прогноз – это интервал значений зависимой переменной, содержащий с вероятностью 0,95 истинное ее значение при заданных значениях независимых переменных. Центр интервала равен точечному прогнозу, концы интервалов получены прибавлением и вычитанием произведения стандартной ошибки прогноза на критическое значение t-статистики, т. е. $y^* \pm t_{\alpha,\nu} S^*$. *Средняя стандартная ошибка прогноза* S^* в матричной форме определяется по формуле

$$S^* = S \cdot \sqrt{X_p^T (X^T X)^{-1} X_p},$$

где S – стандартная ошибка регрессии;

$X_p = \begin{pmatrix} 1 \\ p_1 \\ p_2 \end{pmatrix}$ – матрица заданных для построения прогноза значений независимых переменных $x_1 = p_1$, $x_2 = p_2$;

$X = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} \end{pmatrix}$ – матрица, составленная из столбца n единиц, столбца n значений переменной x_1 и столбца n значений переменной x_2 из исходных данных;

T – индекс, обозначающий операцию транспонирования матрицы.

Поскольку выполняются все условия верификации, то модель является качественной. Следовательно, прогноз, выполненный по ней, является качественным: несмещенным, состоятельным и эффективным. На листе «Регрессия» рассчитан точечный прогноз фактора y, который равен 699,53. На листе «Интервальный прогноз» получен интервальный прогноз (697,38; 701,68), который означает, что с вероятностью 0,95 любое значение из этого интервала является оценкой фактора y.

Вопросы для самоконтроля

1. Каковы этапы построения эконометрической модели?
2. Какова краткая характеристика цели каждого этапа?
3. Знания из каких научных дисциплин необходимы на каждом из этапов эконометрического моделирования?
4. В чем заключается спецификация модели множественной регрессии?
5. Почему в уравнении регрессии присутствует случайная переменная?
6. Как определить силу и направленность взаимодействия факторов?

7. Что означает значимость коэффициента корреляции?

8. Как проверить на значимость коэффициент корреляции?

9. В чем заключается суть МНК для нахождения оценок параметров регрессии?

10. Почему с помощью МНК находятся оценки параметров, а не их точные значения?

11. Какая оценка параметра называется точечной?

12. В чем заключается суть интервальной оценки параметров?

13. Как найти интервальные оценки коэффициентов регрессии?

14. Как используются стандартные ошибки регрессии и стандартные ошибки коэффициентов регрессии при анализе оценок параметров регрессии?

15. В чем заключается экономический смысл параметров моделирегрессии?

16. Как можно оценить общее качество уравнения регрессии?

17. Какова суть коэффициента детерминации, нормированного коэффициента детерминации? В каких пределах они изменяются?

18. Какова связь между коэффициентом детерминации, коэффициентом корреляции и множественным коэффициентом корреляции для множественной регрессии?

19. Для какой цели в парной регрессии используется критерий Фишера?

20. Как получить остатки для модели парной регрессии?

21. Какому условию должны удовлетворять остатки, чтобы для проверки статистических гипотез можно было использовать критерий Стьюдента?

22. Какое распределение называется нормальным? Каковы его параметры?

23. Какими способами можно проверить нормальность распределения остатков?

24. Как используется критерий согласия Пирсона для проверки гипотезы о нормальном законе распределения остатков?

25. Для чего и как проверяется значимость коэффициентов регрессии?

26. Какими свойствами должны обладать оценки параметров регрессии?

27. Каковы основные предпосылки применения МНК для построения регрессионной модели?

28. Каковы последствия невыполнимости предпосылок применения МНК?

29. Как сформулировать теорему Гаусса–Маркова?

30. Как проверить центрированность остатков?

31. В чем заключается суть гетероскедастичности (гомоскедастичности) остатков?

32. Каковы причины и последствия гетероскедастичности остатков?

33. Как определить гомоскедастичность остатков?

34. В чем заключается суть автокорреляции остатков? Каковы ее последствия?

35. Как проверить гипотезу об отсутствии автокорреляции остатков с помощью критерия Дарбина–Уотсона?

36. Как проверить гипотезу об отсутствии автокорреляции остатков визуально?

37. Что характеризуют средние коэффициенты эластичности?

38. Для чего используются частные коэффициенты корреляции? Как они рассчитываются?

39. Каковы последствия мультиколлинеарности остатков?

40. Что представляет собой прогнозирование?

41. Как получить точечный прогноз зависимого фактора?

42. Как получить интервальный прогноз зависимого фактора?

Индивидуальное задание

Выполните следующее:

- исследуйте зависимость фактора y от факторов x_1 и x_2, используя данные наблюдений, приведенные в таблице 1, прибавив к значениям фактора y величину $10 \cdot к$, где $к$ – номер в журнале;
- постройте регрессионную модель $y = f(x_1, x_2) + \varepsilon$;
- рассчитайте значение фактора y, если $x_1 = 35$ и $x_2 = 10$;
- оформите отчет.

Тема 2. ВРЕМЕННЫЕ РЯДЫ

Постановка задачи

Динамика выпуска продукции некоторого предприятия характеризуется данными, представленными в таблице 15.

Таблица 15 – **Исходные данные для анализа структуры временного ряда, млн р.**

Год	Выпуск продукции	Год	Выпуск продукции
2001	5 665	2011	19 037
2002	9 570	2012	21 748
2003	11 172	2013	23 298
2004	10 150	2014	26 570
2005	12 704	2015	26 080
2006	12 588	2016	27 446
2007	13 018	2017	29 658
2008	13 471	2018	32 573
2009	15 017	2019	36 435
2010	17 356	2020	38 100

Проанализируйте структуру временного ряда, проверьте гипотезу о структурной стабильности ряда, проведите аналитическое выравнивание временного ряда, сделайте прогноз на 2021 г.

Технология вычислений в MS Excel при построении модели временного ряда

1. Постановочный этап

Введите подготовленные исходные данные, представленные в таблице 22. В ячейку A1 введите название первого столбца «Год», в ячейку B1 – название второго столбца «Выпуск продукции», в ячейку C1 – название третьего столбца Уровень временного ряда. В ячейки A2, A3, …, A21 введите данные первого столбца исходной таблицы, в ячейки B2, B3,…, B21 – данные второго столбца, в ячейки C2, …, C21 – номера уровней 1, 2, 3, …, 20.

Введите новое название листа «Исходные данные». Сохраните рабочую книгу под названием «Временные ряды».

2. Спецификация: *определение вида аналитической модели временного ряда*

2.1. Анализ структуры временного ряда

2.1.1. Оценка наличия тенденции во временном ряде с помощью корреляционного поля

Для построения корреляционного поля выполните следующие действия:

- выделите диапазон ячеек B1:B21, содержащий данные вместе с названием;
- на вкладке ленты *Вставка* в группе инструментов *Диаграммы* выберите диаграмму типа *Точечная*, а в ней – *Точечная с прямыми отрезками и маркерами*.

2.1.2. Оценка структуры временного ряда: наличие тренда, сезонности, цикличности, случайной компоненты – по автокорреляционной функции временного ряда и коррелограмме

Для обеспечения статистической достоверности коэффициентов автокорреляции определите число периодов, по которым рассчитывается коэффициент автокорреляции по правилу о том, что максимальный лаг должен быть не меньше $n/4 = 20/4 = 5$. На листе «Исходные данные» в ячейку D1 введите название «Период». В ячейки D2:D6 введите значения 1, 2, …, 5 (число периодов τ).

В ячейку E1 введите название «Коэф. автокор.».

В ячейку E2 введите формулу коэффициента корреляции r1

= КОРРЕЛ(B2:B20;B3:B21).

В ячейку E3 введите формулу коэффициента корреляции r2

= КОРРЕЛ(B2:B19;B4:B21).

В ячейку E4 введите формулу коэффициента корреляции r3

= КОРРЕЛ(B2:B18;B5:B21).

В ячейку E5 введите формулу коэффициента корреляции r4

= КОРРЕЛ(B2:B17;B6:B21).

В ячейку E6 введите формулу коэффициента корреляции r5

= КОРРЕЛ(B2:B16;B7:B21).

Для проверки на значимость коэффициентов автокорреляции выполните следующие действия:

- в ячейку G1 введите название «Значимость коэффициентов автокорреляции»;
- в ячейки G2, G3, G4, G5, G6 введите t1, t2, t3, t4, t5 соответственно;
- в ячейку H2 введите формулу = E2*КОРЕНЬ((20 – 2)/(1 – E2^2));
- в ячейки H3:H6 введите формулы, как и в ячейке H2, заменяя ссылки на ячейки и объем выборки 20 всякий раз уменьшая на 1.

В ячейку D7 введите Макс.

В ячейку E7 введите формулу = МАКС(E2:E6).

Вычислите критическое значение статистики следующим образом:

- в ячейку G7 введите tкр;
- в ячейку H7 введите формулу =СТЬЮДЕНТ.ОБР.2Х(0,05;20 – 2).

Если наиболее высоким (первым в списке) окажется коэффициент автокорреляции порядка $\tau > 1$, то ряд содержит сезонные компоненты с периодомτ, тренд и случайную компоненту; если $\tau = 1$, то ряд содержит только тренд и случайную компоненту.

Для построения коррелограммы выполните следующие действия:

- выделите диапазон ячеек D1:E6, содержащий данные вместе с названиями;
- на вкладке ленты *Вставка* в группе инструментов *Диаграммы* выберите диаграмму типа *Точечная*, а в ней – *Точечная с прямыми отрезками и маркерами*.

2.2. Определение вида модели (аддитивная или мультипликативная) по корреляционному полю

Примечание – Подробное описание выбора вида модели приведено в разделе «Эконометрический анализ построения модели временного ряда».

3. Аналитическое выравнивание временного ряда

3.1. Структурная стабильность временного ряда

Определяется по виду корреляционного поля, построенного в подпункте 2.1.1.

3.2. Проведение аналитического выравнивания временного ряда

Для аналитического выравнивания временного ряда выполните следующие действия:

- выделите диапазон ячеек A1:B21, содержащий данные вместе с названиями;

• на вкладке ленты *Вставка* в группе инструментов *Диаграммы* выберите диаграмму типа *Точечная*, а в ней – *Точечная с прямыми отрезками и маркерами*.

Выделите полученную кривую нажатием левой кнопки мыши, затем на ней же нажмите правую кнопку мыши и выберите *Добавить линию тренда*. В диалоговом окне выберите линейный тип линии тренда, установите флажки в полях *показывать уравнение на диаграмме* и *поместить на диаграмму величину достоверности аппроксимации (R*^2).

Аналогично постройте экспоненциальную, логарифмическую, полиномиальные степени 2 и 3, степенную линии тренда, каждый раз попарно сравнивая R^2, удаляйте линию с наименьшим показателем R^2.

4. Верификация

В результате предыдущих действий останется линия тренда с максимальным R^2. Соответствующее ей уравнение будет наилучшей формой тренда.

5. Прогнозирование

В ячейку A23 введите год 2021. В ячейке B23 рассчитайте для выбранной формы прогнозируемое значение выпуска продукции по формуле трендовой модели при $x = 22$ (порядковый номер года – 2021, считая 2001 г. первым по порядку).

Эконометрический анализ построения модели временного ряда

1. Постановочный этап

> Временной ряд – это совокупность значений какого-либо показателя за ряд последовательных периодов времени.
>
> Модель, построенная по данным временного ряда, называется моделью временного ряда.
>
> Выделяют следующие основные задачи исследования временных рядов:
>
> • характеристику структуры временных рядов;
>
> • определение вида аналитической модели временного ряда;
>
> • определение аналитической модели временного ряда;
>
> • анализ качества модели;
>
> • прогнозирование по модели временного ряда.

Динамика выпуска продукции представлена в таблице 15 наблюдениями, характеризующими некоторое предприятие за 20 последовательных лет. Возникает задача количественного описания динамики выпуска продукции с использованием модели временного ряда, где t – независимая переменная, y – зависимая переменная.

2. Спецификация: *определение вида аналитической модели временного ряда*

2.1. Анализ структуры временного ряда

Уровни временного ряда – это значения $y_t (t = 1, 2, ..., n)$ наблюдаемого показателя в каждом из n временных периодов (дней, недель, месяцев, кварталов, лет).

Каждый уровень формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

- факторы, формирующие тренд ряда, т. е. изменение динамики значений исследуемого показателя под совместным долговременным воздействием множества факторов (рисунок 11 а);
- факторы, формирующие циклические (периодические колебания, выходящие за рамки более короткого периода, например, одного года) и сезонные колебания (периодические колебания в рамках более короткого периода, например, года) (рисунок 11 б);
- случайные факторы, которые нерегулярно воздействуют на временной ряд (например, факторы резкого и внезапного действия), ошибки наблюдений (рисунок 11 в).

Под воздействием этих факторов проявляется тенденция изменения (возрастания, убывания, стабильности) временного ряда, которая содержит все три компоненты: случайную ошибку, тренд, сезонность (цикличность). Некоторые временные ряды не содержат тренда и циклическую компоненту (рисунок 11 (в)), а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой случайной компоненты. Такие временные ряды называются стационарными. Для них такие статистические характеристики временного ряда как его математическое ожидание (среднее), дисперсия (ср. кв. отклонение) и ковариация не зависят от времени. Если нарушается хотя бы одно из этих условий, то ряд называется нестационарным (рисунок 11 (а) и 11 (б)).

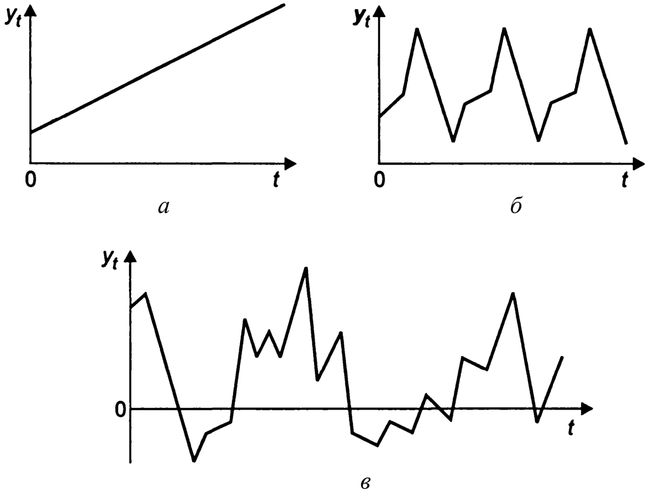


Рисунок 11 –**Основные компоненты временного ряда: линейный тренд, возрастающая тенденция (*а*); только сезонная компонента (*б*); только случайная компонента (*в*)**

Уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной структурных компонент ряда.

Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется аддитивной моделью временного ряда.

Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется мультипликативной моделью временного ряда.

Выбор одной из двух моделей проводится на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

По реальным данным строится модель, содержащая хотя бы одну компоненту.

Основная задача эконометрического исследования отдельного временного ряда заключается в выявлении и придании количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

2.1.1. Оценка наличия тенденции во временном ряде с помощью корреляционного поля

Корреляционным полем временного ряда называется множество точек на плоскости с координатами (t; y). По виду корреляционного поля оценивается тенденция, характеризующая совокупное долговременное воздействие множества факторов на динамику изучаемого показателя, которые в совокупности формируют его возрастание или убывание либо их отсутствие.

По виду корреляционного поля временной ряд содержит возрастающую тенденцию (рисунок 12). Временной ряд, очевидно, содержит трендовую компоненту (предположительно линейный тренд), поэтому визуально является нестационарным.

Рисунок 12 – **Корреляционное поле временного ряда**

2.1.2. Оценка структуры временного ряда: наличие тренда, сезонности, цикличности, случайной компоненты – по автокорреляционной функции временного ряда и коррелограмме

При наличии тенденции значение каждого последующего уровня ряда зависит от предыдущих значений. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями y_t исходного временного ряда и уровнями $y_{t-\tau}$ этого ряда, сдвинутыми на τ шагов во времени, по следующей формуле:

$$r_\tau = \frac{\sum_{t=1}^{n-\tau}\left(y_t - \bar{y}_1\right)\cdot\left(y_{t+\tau} - \bar{y}_2\right)}{\sqrt{\sum_{t=1}^{n-\tau}\left(y_t - \bar{y}_1\right)^2 \cdot \sum_{t=\tau+1}^{n}\left(y_{t+\tau} - \bar{y}_2\right)^2}},$$

где $\bar{y}_1 = \frac{\sum_{t=1}^{n-\tau} y_t}{n-\tau}$;

$$\bar{y}_2 = \frac{\sum_{t=\tau+1}^{n} y_t}{n-\tau}.$$

Число периодов τ, по которым рассчитывается коэффициент автокорреляции, называется лагом. Считается целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило: максимальный лаг должен быть не больше n/4.

Коэффициент автокорреляции является аналогом линейного коэффициента корреляции, изменяется в пределах от –1 до 1. Его значимость оценивается с помощью t-статистики: $t = r_\tau\sqrt{(n-2)/\left(1-r_\tau^2\right)}$, которая имеет распределение Стьюдента с $k = n-2$ степенями свободы. Если $\left|t\right| \geq t_{\alpha,k}$, то проверяемый коэффициент значим.

На основании автокорреляционных коэффициентов осуществляется анализ структуры временного ряда. Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только трендовую и случайную компоненты. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, то ряд содержит циклические колебания с периодичностью τ. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать предположение относительно структуры этого ряда: либо ряд не содержит тренда и циклических колебаний, а только случайную компоненту, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

По коэффициенту автокорреляции можно судить о наличии линейного тренда. Если временной ряд имеет сильную нелинейную тенденцию, то коэффициент автокорреляции уровней исходного ряда может приближаться к нулю. По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда.

Последовательность коэффициентов автокорреляции уровней первого, второго и следующих порядков называют *автокорреляционной функцией временного ряда.* График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой.* При помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Поскольку максимальным является коэффициент автокорреляции первого порядка (таблица 16 и рисунок 13), то исследуемый ряд не содержит сезонной компоненты, т. е. содержит только трендовую и случайную структурные компоненты.

Таблица 16 – **Значения автокорреляционной функции**

Период	Коэффициент автокорреляции
1	0,99
2	0,98
3	0,98
4	0,97
5	0,97

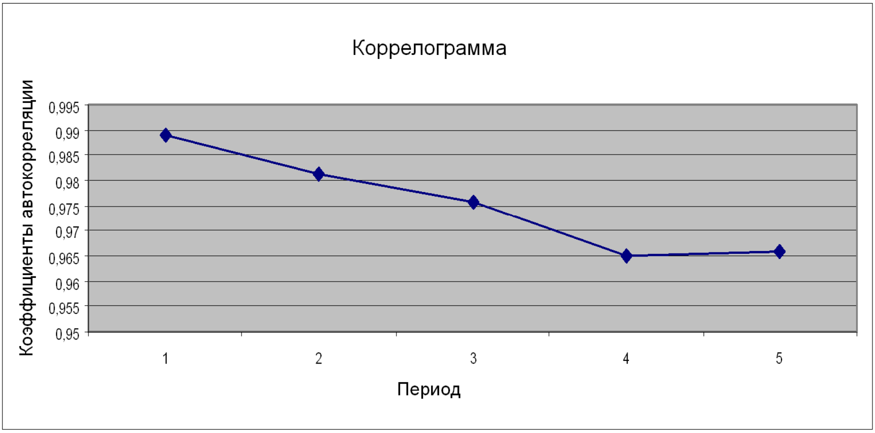


Рисунок 13 – **Автокорреляционная функция**

Поскольку все наблюдаемые значения *t*-статистики больше критического (таблица 17), то все коэффициенты автокорреляции значимы. Это подтверждает предположение о том, что ряд не содержит сезонных и циклических компонент, а содержит только трендовую и случайную компоненты.

Таблица 17 – **Значимость коэффициентов автокорреляции**

Значимость	
t1	28,15
t2	21,57
t3	18,97
t4	15,66
t5	15,80
tкр	2,10

2.2. Определение вида модели (аддитивная или мультипликативная) по корреляционному полю

Если имеются несколько компонент, то необходимо определить, как объединить их в модели: умножением или сложением. Для этого анализируется корреляционное поле. Если амплитуда колебаний временного ряда затухает, то временной ряд описывается мультипликативной моделью, если амплитуда колебаний постоянна, то – аддитивной моделью.

Анализ рисунка 12 показывает, что колебания временного ряда отсутствуют. Следовательно, модель аддитивна и представляет собой сумму трендовой и случайной компоненты.

3. Аналитическое выравнивание временного ряда

3.1. Структурная стабильность временного ряда

Существуют единовременные изменения характера тенденции временного ряда, вызванные структурными изменениями в экономике или иными факторами. Они отличаются от сезонных и циклических колебаний тем, что, начиная с некоторого момента времени, происходит изменение характера динамики изучаемого показателя. Это приводит к изменению параметров тренда, описывающего эту динамику (рисунок 14).

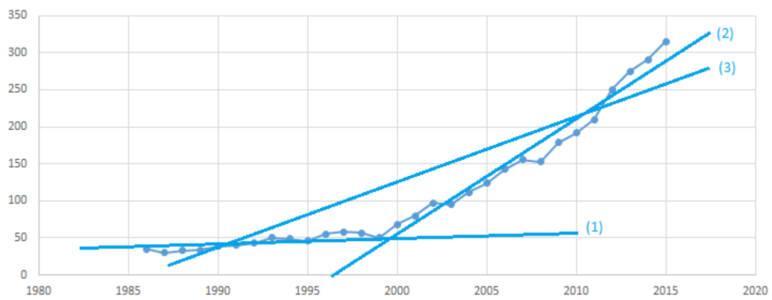


Рисунок 14 – **Изменение характера тенденции временного ряда**

В момент времени t^* значительно изменяется влияние ряда факторов на изучаемый показатель (например, изменения структуры экономики: начало крупных экономических реформ, изменение экономического курса, нефтяные кризисы и прочие факторы). Возникает вопрос о том, значимо ли повлияли общие структурные изменения на характер тенденции изменения изучаемого показателя.

Если это влияние значимо, то для моделирования тенденции данного временного ряда следует использовать кусочно-линейные модели регрессии, т. е. разделить исходную совокупность на две подсовокупности (до момента времени t^* и после него) и построить отдельно по каждой подсовокупности уравнения линейной регрессии (на рисунке 14 этим уравнениям соответствуют прямые (1) и (2). Если структурные изменения незначительно влияли на характер тенденции ряда, то ее можно описать с помощью единого для всей совокупности данных уравнения (на рисунке 14 этому уравнению соответствует прямая (3).

Остаточная сумма квадратов кусочно-линейной модели меньше по сравнению с общим для всей совокупности уравнением тренда. Однако разделение исходной совокупности на две части ведет к потере числа наблюдений и, следовательно, снижению числа степеней свободы в каждом уравнении кусочно-линейной модели. Построение единого для всей совокупности уравнения тренда, напротив, позволяет сохранить число наблюдений n исходной совокупности.

Выбор одной из двух моделей (кусочно-линейной или единого уравнения тренда) будет зависеть от соотношения между снижением остаточной дисперсии и потерей числа степеней свободы при переходе от единого уравнения регрессии к кусочно-линейной модели.

Гипотеза о наличии структурной стабильности временного ряда проверяется по тесту Чоу при справедливости гипотезы о равенстве генеральных дисперсий для выделенных подвыборок.

Пусть S_0 – сумма квадратов отклонений выборочных значений от модельных общего уравнения регрессии (3), S_1 – сумма квадратов отклонений выборочных значений от модельных уравнения регрессии по первой подвыборке (1), S_2 – по второй подвыборке (2). Равенство $S_0 = S_1 + S_2$ возможно лишь при совпадении коэффициентов регрессии для всех трех уравнений. Чем сильнее различие в поведении зависимой переменой для двух подвыборок, тем больше значение S_0 будет превосходить $S_1 + S_2$.

Основная гипотеза формулируется как утверждение о том, что качество общей модели регрессии лучше качества моделей регрессии подвыборок. Альтернативная гипотеза утверждает, что качество общей модели регрессии хуже качества моделей регрессии подвыборок. Для проверки данной гипотезы используется F-статистика, имеющая распределение Фишера и рассчитываемая по формуле:

$$F = \frac{S_0 - S_1 - S_2}{S_1 + S_2} \cdot \frac{n - 2m - 2}{m + 1}, \tag{4}$$

где n – объем выборки, m – количество независимых переменных в уравнениях регрессии. Критическое значение F-критерия определяется по таблице распределения Фишера–Снедекора в зависимости от уровня значимости α и двух степеней свободы $k_1 = m + 1$ и $k_2 = n - 2m - m$.

По корреляционному полю на рисунке 12 определяем, что изменения характера динамики временного ряда не наблюдается, так как точки расположены вдоль одной линии. Значит нет необходимости использовать кусочно-линейные модели регрессии.

3.2. Проведение аналитического выравнивания временного ряда

Одним из наиболее распространенных способов моделирования временного ряда является построение тренда или аналитической функции, характеризующей зависимость уровней ряда от времени. Этот способ называют *аналитическим выравниванием временного ряда*.

Для аналитического выравнивания могут применяться следующие функции:

- линейная $\tilde{y}_t = a + b \cdot t$;
- гиперболическая $\tilde{y}_t = a + \dfrac{b}{t}$;
- экспоненциальная $\tilde{y}_t = e^{a+b\cdot t}$;
- степенная $\tilde{y}_t = a \cdot t^b$;
- полиномы второго и более высоких порядков $\tilde{y}_t = a + b_1 \cdot t + + b_2 \cdot t^2 + \ldots + b_k \cdot t^k$.

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время, а в качестве зависимой переменной – фактические уровни временного ряда yt. Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Известно несколько способов определения типа трендов. К наиболее распространенным относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени, расчет некоторых основных показателей динамики, коэффициенты автокорреляции уровней ряда. Тип тренда можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет линейный тренд, то его соседние уровни тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит нелинейный тренд, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражен нелинейный тренд в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

4. Верификация

Выбор наилучшего уравнения в случае, если ряд содержит нелинейный тренд, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации R_2, значимость которого оценивается по критерию Фишера, и выбора уравнения тренда с максимальным значением скорректированного коэффициента детерминации. Реализация этого метода относительно проста при компьютерной обработке данных.

При наличии неявного нелинейного тренда следует дополнять описанные выше методы выбора наилучшего уравнения тренда качественным анализом динамики изучаемого показателя, чтобы избежать ошибок спецификации при выборе вида тренда. Качественный анализ предполагает изучение проблем возможного наличия в исследуемом временном ряде поворотных точек и изменения темпов прироста, начиная с определенного момента (периода) времени под влиянием ряда факторов. В случае, если уравнение тренда выбрано неверно при больших значениях выборки (ошибка спецификации), результаты анализа и прогнозирования динамики временного ряда с использованием выбранного уравнения будут недостоверными.

Поскольку наибольшее значение коэффициента детерминации 0,98 имеет уравнение, заданное кубическим полиномом, то в качестве модели можно использовать это уравнение (рисунок 15). Однако значение коэффициента детерминации линейного тренда равно 0,96, что также дает право использовать его для прогноза. Как правило, при прогнозировании предпочтение отдается линейному тренду, если по качеству он незначительно уступает нелинейному.

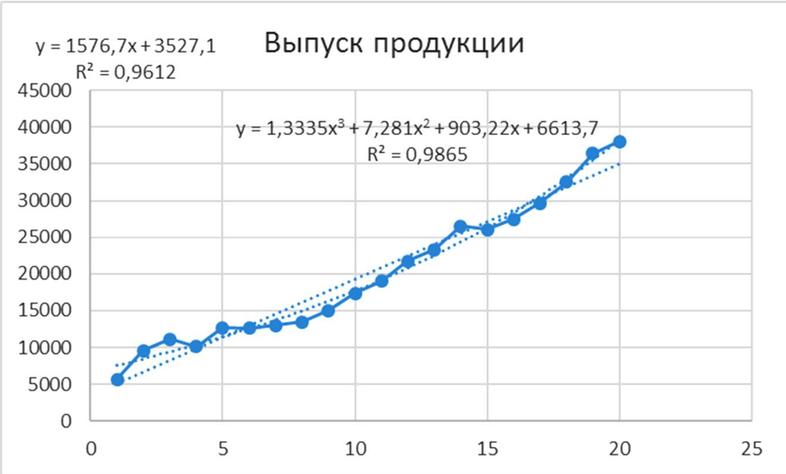


Рисунок 15 – **Подбор линии тренда**

5. Прогнозирование

Используя линию тренда (кубический полином), осуществляется прогноз выпуска продукции, который составит в 2021 г. 44 208 ед. Прогноз выпуска продукции по линейному тренду составит 38 214,5 ед. Заметим, что полином лучше описывает имеющуюся выборку, но прогнозное значение резко увеличивается по сравнению с наблюдаемыми значениями. Прогноз по линейному тренду более достоверен.

Вопросы для самоконтроля

1. Каково определение модели временного ряда?
2. Какие известны основные компоненты временного ряда?
3. Каковы основные цели исследования временных рядов?

4. Как использовать автокорреляционную функцию при анализе структуры временного ряда?

5. Как рассчитывается коэффициент автокорреляции пятого порядка?

6. Как строится коррелограмма?

7. Каков общий вид мультипликативной и аддитивной моделей временного ряда?

8. С какой целью проводится анализ структуры сезонных колебаний временного ряда?

9. Какие тесты используются для проверки гипотезы о структурной стабильности временного ряда?

10. В каком случае нарушается структурная стабильность временного ряда?

11. Что понимается под аналитическим выравниванием временного ряда?

12. Каковы известны наиболее распространенные модели, используемые для аналитического выравнивания временного ряда?

13. Что понимается под линеаризующими преобразованиями? Как они используются в МНК?

14. Как оценивается качество построенной модели?

15. Как осуществляется точечный прогноз по модели временного ряда?

Индивидуальное задание

Динамика выпуска продукции некоторого предприятия характеризуется данными, представленными в таблице 15 (в каждом варианте к объему выпускаемой продукции надо прибавить число $120 \cdot k$, где k – порядковый номер студента в журнале группы). Выполните следующее:

- проанализируйте структуру временного ряда;
- проверьте гипотезу о структурной стабильности ряда;
- проведите аналитическое выравнивание временного ряда;
- сделайте прогноз на 2021 г.;
- оформите отчет.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Аистов, А. В. Эконометрика шаг за шагом : учеб. пособие / А. В. Аистов, А. Г. Максимов. – М. : ГУ ВШЭ, 2006. – 177 с.

Айвазян, С. А. Прикладная статистика и основы эконометрики : учеб. / С. А. Айвазян, В. С. Мхитрян. – М. : ЮНИТИ-Дана, 1998. – 1022 с.

Айвазян, С. А. Прикладная статистика. Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1985. – 488 с.

Айвазян, С. А. Прикладная статистика. Основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1983. – 471 с.

Алексеенко, В. Б. Математические методы исследования экономических систем : учеб. пособие / В. Б. Алексеенко, В. В. Красавина. – М. : РУДН, 2005. – 154 с.

Бородич, С. А. Эконометрика : учеб. пособие / С. А. Бородич. – Минск : Новое знание, 2001. – 408 с.

Грубер, Й. Эконометрия : учеб. пособие. В 2 т. Т. 2 : Эконометрические прогнозные и оптимизационные модели / Й. Грубер. – Киев : Нічлава, 1999. – 308 с.

Доугерти, К. Введение в эконометрику : учеб. пособие : [пер. с англ.] / К. Доугерти; под ред. О. О. Замкова. – М. : ИНФРА-М, 1997. – 402 с.

Елисеева, И. И. Эконометрика : учеб. / И. И. Елисеева. – М. : Финансы и статистика, 2006. – 576 с.

Карминский, А. М. Рейтинги в экономике: методология и практика : учеб. пособие / А. М. Карминский. – М. : Финансы и статистика, 2005. – 240 с.

Красс, М. С. Математические методы и модели для магистрантов и экономистов : учеб. пособие / М. С. Красс, Б. П. Чупрынов. – СПб. : Питер, 2006. – 496 с.

Эконометрика и экономико-математические методы и модели : учеб. пособие / под ред. Г. О. Читая, С. Ф. Миксюк. – Минск : БГЭУ, 2018. – 511 с.

Магнус, Я. Р. Эконометрика. Начальный курс : учеб. пособие / Я. Р. Магнус. – М. : Дело, 1997. – 248 с.

Маленво, Э. Статистические методы эконометрии / Э. Маленво. – М. : Статистика, 1976. – 329 с.

Мардас, А. Н. Эконометрика : учеб. пособие / А. Н. Мардас. – СПб. : Питер, 2001. – 144 с.

Марченко, Л. Н. Теория вероятностей и математическая статистика : практ. пособие / Л. Н. Марченко, Л. П. Авдашкова. – Гомель : ГГУ им. Ф. Скорины, 2007. – 275 с.

Нименья, И. Н.Эконометрика : учеб. пособие / И. Н. Нименья. – СПб. : Нева, 2004. – 224 с.

Новак, Э. Введение в методы эконометрики : сб. задач / Э. Новак ; пер. с пол. под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2006. – 248 с.

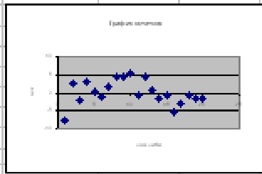
ПРИЛОЖЕНИЕ

Результаты расчетов по теме «Множественная регрессия»

	A	B	C	D	E	F	G	H
1	Фактор y	Фактор x1	Фактор x2		Корреляционная матрица			
2	684,48	28	10			*Фактор y*	*Фактор x1*	*Фактор x2*
3	674,45	26	8		Фактор y	1		
4	729,62	30	14		Фактор x1	0,794825	1	0,747928
5	748,86	35	15		Фактор x2	0,995338	0,747928	1
6	761,44	41	16		Значимость коэффициентов корреляции			
7	773,42	45	17		tнабл y,x1	5,556924		
8	628,07	27	3		tнабл y,x2	43,7846		
9	731,84	35	13		tкр.	2,100922		
10	698,81	30	10					
11	645,92	23	5		Мультиколлинеарность			
12	664,64	29	7		Определитель	0,440604		
13	711,18	33	11		хи-кв набл.	19,53393		
14	798,07	40	20		хи-кв кр.	223,1602		
15	833,82	41	24					
16	667,97	41	6		Частные коэффициенты корреляции			
17	607,61	23	2		r y,x1-x2	0,787012		
18	711,76	32	12		r y,x2-x1	0,995181		
19	728,62	37	13		tнабл (r y,x1-x2)	5,412226		
20	666,15	31	7		tнабл (r y,x2-x1)	43,06106		
21	683,7	30	9					
22								
23	Эластичность							
24	Фактор y_ср	Фактор x1_ср	Фактор x2_ср					
25	707,5215	32,85	11,1					
26	Коэф. фактора x1	1,029554971						
27	Эластичность фактора x1	0,047801912						
28	Коэф. фактора x2	9,275783402						
29	Эластичность фактора x2	0,14552377						

Рисунок 1 – **Лист «Исходные данные»**

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ				Точечный прогноз				
2					699,5316814				
3	рессионная статистика								
4	Множестве	0,998228							
5	R-квадрат	0,99646							
6	Нормиров:	0,996043							
7	Стандартн	3,597326							
8	Наблюдени	20							
9									
10	Дисперсионный анализ								
11		df	SS	MS	F	Значимость F			
12	Регрессия	2	61917,595	30958,79731	2392,348568	1,4688E-21			
13	Остаток	17	219,99284	12,94075525					
14	Итого	19	62137,587						
15				Fкр.	3,591530569				
16		Коэффициенты	ндартная ош	t-статистика	P-Значение	Нижние 95%	Верхние 95%	ижние 95,0	рхние 95,0%
17	Y-пересеч	570,7394	4,9378112	115,5855106	4,62747E-26	560,321553	581,157294	560,3216	581,1573
18	Фактор x1	1,029555	0,1957426	5,259738186	6,38714E-05	0,61657413	1,44253581	0,616574	1,442536
19	Фактор x2	9,275783	0,2216551	41,8478256	1,37064E-18	8,80813205	9,74343475	8,808132	9,743435
20			tкр	2,109815559					
21									
22									
23	ВЫВОД ОСТАТКА								
24									
25	Наблюдение	редсказанное	Остатки	Условие 1		Условие 3			
26	1	692,3248	-7,8447966	Остатки		111,95072	61,5408344		
27	2	671,7141	2,7358801			21,1868753	7,48503993		
28	3	731,487	-1,8670402	Среднее	-2,84217E-14	23,198111	3,4858391		
29	4	745,9106	2,9494015	Стандартная	0,760873527	8,25477964	8,69896946		
30	5	761,3637	0,0762883	Медиана	-0,717009631	1,9994053	0,00581991		
31	6	774,7577	-1,337715	Мода	#Н/Д	9,25958803	1,78948136		
32	7	626,3648	1,7052421	Стандартное	3,402729859	7,70465596	2,90785076		
33	8	727,359	4,4809683	Дисперсия в	11,57857049	0,00301126	20,0790773		
34	9	694,3839	4,4260934	Эксцесс	0,102240561	0,48414016	19,5903029		
35	10	640,7981	5,1218952	Асимметричн	-0,360269088	36,1068388	26,2338107		
36	11	665,527	-0,8870014	Интервал	12,96669187	28,2880006	0,7867715		
37	12	706,7484	4,4316451	Минимум	-7,844796646	14,4319103	19,6394782		
38	13	797,4373	0,6327097	Максимум	5,121895223	5,67720487	0,40032153		
39	14	835,57	-1,7499789	Сумма	-5,68434E-13	1,24122158	3,06242618		
40	15	668,6059	-0,6358777	Счет	20	22,3244618	0,4043404		
41	16	612,9708	-5,3607546	tнабл	-3,73541E-14	4,52060411	28,7376896		
42	17	714,9946	-3,2345833	tкр	2,09302405	5,93624835	10,4625294		
43	18	729,4181	-0,7981416			0,40700541	0,63703001		
44	19	667,5861	-1,4361114			0,00078334	2,06241582		
45	20	685,1081	-1,4081232				1,98281091		
46						302,975566	219,992839		
47						d набл	1,37720649		
48									
49	Критерий Пирсона								
50	Карман	Частота	0						
51	-7,844797	1	0,0105709	0,010570913	0,211418264	2,94137859			
52	-4,603124	1	0,088064	0,077493045	1,549860904	0,1950801			
53	-1,361451	5	0,3445394	0,256475472	5,129509443	0,00326984			
54	1,880222	7	0,7097186	0,365179191	7,303583821	0,01261889			
55	15,36569	6	0,9999968	0,290278223	5,805564458	0,00651189			
56									
57	хи кв наб	3,158859							
58	хи кв кр	7,814728							



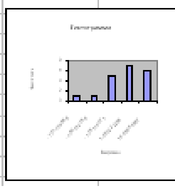
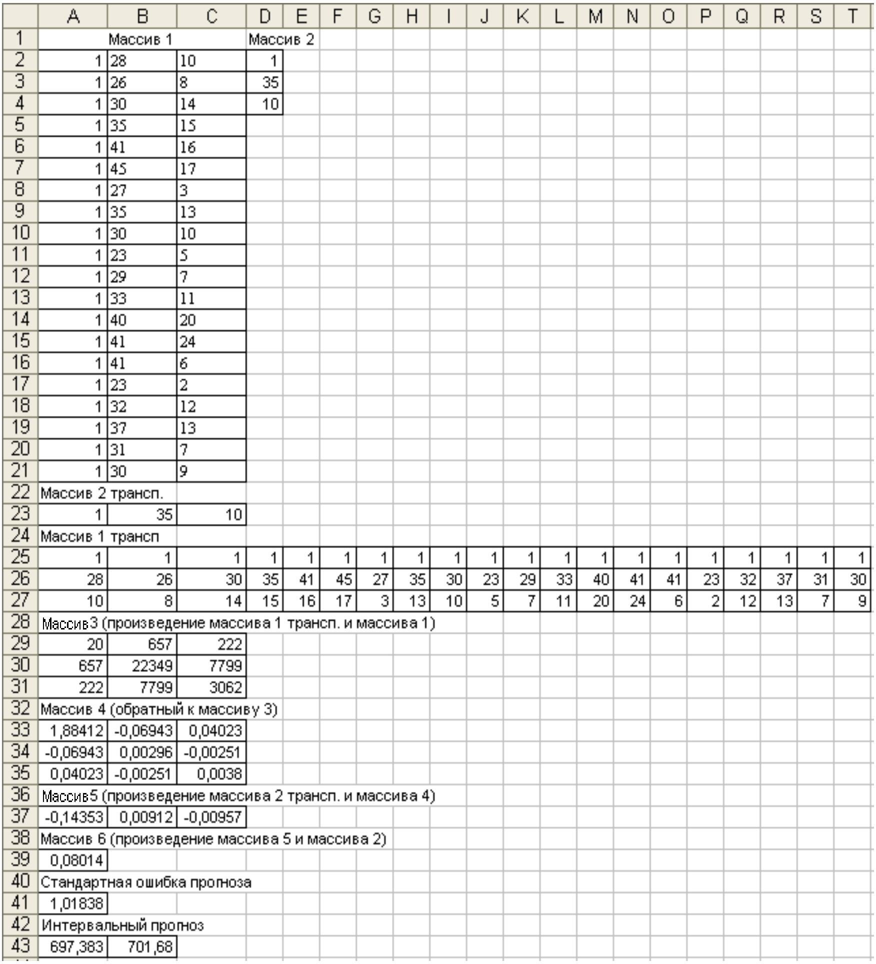


Рисунок 2 – **Лист «Регрессия»**

	A	B	C	D	E	F	G	H	I	J	K	L
1	Фактор y	*Остатки*	Мод ост.	*Точка*	Фактор y	*Ранг*	*Процент*	*Точка*	*Мод ост.*	*Ранг*	*Процент*	*Квадрат разности рангов*
2	684,48	-7,8447966	7,844797	1	684,48	12	42,10%	1	7,844797	1	100,00%	121
3	674,45	2,7358801	2,73588	2	674,45	14	31,50%	2	2,73588	9	57,80%	25
4	729,62	-1,8670402	1,86704	3	729,62	7	68,40%	3	1,86704	10	52,60%	9
5	748,86	2,94940154	2,949402	4	748,86	5	78,90%	4	2,949402	8	63,10%	9
6	761,44	0,07628831	0,076288	5	761,44	4	84,20%	5	0,076288	20	0,00%	256
7	773,42	-1,337715	1,337715	6	773,42	3	89,40%	6	1,337715	15	26,30%	144
8	628,07	1,70524214	1,705242	7	628,07	19	5,20%	7	1,705242	12	42,10%	49
9	731,84	4,48096835	4,480968	8	731,84	6	73,60%	8	4,480968	4	84,20%	4
10	698,81	4,42609341	4,426093	9	698,81	11	47,30%	9	4,426093	6	73,60%	25
11	645,92	5,12189522	5,121895	10	645,92	18	10,50%	10	5,121895	3	89,40%	225
12	664,64	-0,8870014	0,887001	11	664,64	17	15,70%	11	0,887001	16	21,00%	1
13	711,18	4,43164509	4,431645	12	711,18	10	52,60%	12	4,431645	5	78,90%	25
14	798,07	0,63270967	0,63271	13	798,07	2	94,70%	13	0,63271	19	5,20%	289
15	833,82	-1,7499789	1,749979	14	833,82	1	100,00%	14	1,749979	11	47,30%	100
16	667,97	-0,6358777	0,635878	15	667,97	15	26,30%	15	0,635878	18	10,50%	9
17	607,61	-5,3607546	5,360755	16	607,61	20	0,00%	16	5,360755	2	94,70%	324
18	711,76	-3,2345833	3,234583	17	711,76	9	57,80%	17	3,234583	7	68,40%	4
19	728,62	-0,7981416	0,798142	18	728,62	8	63,10%	18	0,798142	17	15,70%	81
20	666,15	-1,4361114	1,436111	19	666,15	16	21,00%	19	1,436111	13	36,80%	9
21	683,7	-1,4081232	1,408123	20	683,7	13	36,80%	20	1,408123	14	31,50%	1
22											tнабл.	-1,2453997
23											tкр.	2,10092204

Рисунок 3 – **Лист «Условие2»**



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		Массив 1		Массив 2
2	1	28	10	1
3	1	26	8	35
4	1	30	14	10
5	1	35	15
6	1	41	16
7	1	45	17
8	1	27	3
9	1	35	13
10	1	30	10
11	1	23	5
12	1	29	7
13	1	33	11
14	1	40	20
15	1	41	24
16	1	41	6
17	1	23	2
18	1	32	12
19	1	37	13
20	1	31	7
21	1	30	9
22	Массив 2 трансп.
23	1	35	10
24	Массив 1 трансп
25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	28	26	30	35	41	45	27	35	30	23	29	33	40	41	41	23	32	37	31	30
27	10	8	14	15	16	17	3	13	10	5	7	11	20	24	6	2	12	13	7	9
28	Массив3 (произведение массива 1 трансп. и массива 1)
29	20	657	222
30	657	22349	7799
31	222	7799	3062
32	Массив 4 (обратный к массиву 3)
33	1,88412	-0,06943	0,04023
34	-0,06943	0,00296	-0,00251
35	0,04023	-0,00251	0,0038
36	Массив5 (произведение массива 2 трансп. и массива 4)
37	-0,14353	0,00912	-0,00957
38	Массив 6 (произведение массива 5 и массива 2)
39	0,08014
40	Стандартная ошибка прогноза
41	1,01838
42	Интервальный прогноз
43	697,383	701,68

Рисунок 4 – **Лист «Интервальный прогноз»**

СОДЕРЖАНИЕ

Учебное издание

ЭКОНОМЕТРИКА И ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ

Пособие
для реализации содержания образовательных программ высшего образования I ступени и переподготовки руководящих работников и специалистов

В трех частях

Часть 3

Авторы-составители:
Авдашкова Людмила Павловна
Грибовская Марал Атаевна

Редактор Т. В. Гавриленко
Компьютерная верстка Л. Ф. Барановская

Подписано в печать 07.05.21. Формат $60 \times 84\ ^1/_{16}$.
Бумага офсетная. Гарнитура Таймс. Ризография.
Усл. печ. л. 3,95. Уч.-изд. л. 3,67. Тираж 100 экз.
Заказ №

Издатель и полиграфическое исполнение:
учреждение образования «Белорусский торгово-экономический университет потребительской кооперации».
Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий
№ 1/138 от 08.01.2014.
Просп. Октября, 50, 246029, Гомель.
http://www.i-bteu.by